hilary putnam

A obținut licența în matematică și filosofie la Universitatea Pennsylvania în anul 1948;

Devine doctor în filosofie la UCLA, sub îndrumarea lui H. Reichenbach, cu o disertație despre conceptul de probabilitate;

A predat matematica la Universitatea de Nord-Vest și la Princeton, filosofia științei la MIT iar în anul 1976 vine la Harvard, unde va rămâne până la retragere (2000) în cadrul departamentului de filosofie;

Este unul dintre cei mai importanți filosofi ai secolului trecut, cu importante contribuții în filosofia limbajului, filosofia minții, filosofia științei și metafilosofie.

hilary putnam

raţiune şi istorie

traducere din limba engleză:
ionel narița

EDITURA TEHNICA 55

Bucureşti, 2005

Ediția originală:
Hilary Putnam - **Reason, Truth and History**
Published by the Press Syndicate of the University of Cambridge
The Pitt Building Trumpington Street, Cambridge Cb2 1RP, UK.

Copyright © 1981 Cambridge University Press
All rights reserved.
The Edinburgh Building, Shaftesbury Road, Cambridge Cb2 2RU, UK.

Ediția în limba română:
Copyright © 2005 S.C. Editura TEHNICĂ S.A.
Toate drepturile asupra acestei ediții sunt rezervate editurii.

Adresă: **S.C. Editura TEHNICĂ S.A.**
Str. Olari, nr. 23, sector 2
cod 024056
Bucureşti, România

www.tehnica.ro

coordonatorul colecției
roman**chirilă**

coperta colecției
floriana**bălan**

coordonator tehnic
florin**gealapu**

design copertă & pre-press
cătălina**măgureanu**

procesare pc
mariana**gheorghiță**

Descrierea CIP a Bibliotecii Naționale a României
PUTNAM, HILARY
Rațiune, adevăr şi istorie / Hilary Putnam ; trad.: Ionel Narița. - Bucureşti : Editura Tehnică, 2005

ISBN 973-31-2259-9

I. Narița, Ionel (trad.)

1

Prefaţă

În această lucrare am intenţia să abordez câteva dihotomii care preocupă atât filosoful, cât şi juristul. Dintre acestea, mai importantă este dihotomia dintre obiectiv şi subiectiv privind adevărul şi raţiunea. O dihotomie, precum cea dintre 'obiectiv' şi 'subiectiv', este acceptată nu atât ca o simplă pereche de categorii, cât ca o caracterizare a unor perspective sau de stiluri de gândire, iar gânditorii se raportează la ele ca la nişte etichete ideologice.

Mulţi filosofi, poate cei mai mulţi, susţin diferite versiuni ale teoriei adevărului 'copie', respectiv, concepţia potrivit căreia o judecată este adevărată numai atunci când 'corespunde faptelor (independente de mental)'; filosofii de această factură consideră că singura alternativă este negarea obiectivităţii adevărului şi capitularea în faţa ideii că toate schemele de gândire şi

toate punctele de vedere sunt fără speranţă subiective. Inevitabil, o minoritate curajoasă (Kuhn, cel puţin în unele lucrări; Feyerabend şi alţi filosofi continentali cum ar fi Foucault) se înscriu sub eticheta opusă. Ei sunt de acord *că alternativa faţă de teoria naivă a adevărului copie înseamnă a privi sistemele de gândire, ideologiile, ba chiar (în cazul lui Kunh şi Feyerabend) teoriile ştiinţifice drept subiective, aşa încât, ei merg mai departe,* dezvoltând *cu vigoare un punct de vedere relativist şi subiectiv.*

Deşi această dispută filosofică are caracterul disputelor ideologic,e nu este, prin ea însăşi, în mod automat rea*: ideile noi, chiar în ştiinţele cele mai exacte sunt în mod frecvent înfierate cu ardoare partizană. În politică bunăoară, polarizarea şi fervoarea ideologică sunt uneori necesare pentru a aduce seriozitate într-o dispută. Dar, odată cu timpul, atât în filosofie, cât şi în politică, ideile noi ajung să fie vechi; ceea ce într-o vreme era provocator, devine predictibil şi plictisitor; iar ideile care odată erau în centrul dezbaterilor, mai târziu sunt menţinute în discuţie numai pentru a sublinia noile alternative. Acelaşi lucru se întâmplă acum şi în ce priveşte dezbaterea între perspectiva adevărului corespondenţă şi perspectivele subiectiviste. În primele trei capitole ale acestei cărţi voi încerca să dezvolt o concepţie asupra adevărului care să reunească cele două aspecte, subiectiv şi obiectiv. Această viziune, cel puţin în intenţie, se inspiră din ideile lui Immanuel Kant, susţinând că teoria naivă a adevărului corespondenţă poate fi respinsă fără a susţine că o asemenea opţiune este o problemă de Zeitgeist, sau o problemă de 'schimbare a gestaltului', sau o problemă de ideologie.*

Punctul de vedere pe care îl apăr, în linii mari, este că există o strânsă conexiune între adevăr *şi* raţionalitate*; că, pentru a o spune mai direct, singurul criteriu pentru ce este un fapt este* raţionalitatea *sa. (Înţeleg aceasta aproape literal şi direct; astfel, dacă se poate accepta raţional că un tablou este frumos, atunci tabloul este frumos devine un* fapt.*) În această perspectivă pot exista* fapte de valoare. *Dar relaţia între acceptabilitate raţională*

şi adevăr este o relaţie între două noţiuni distincte. O judecată ar putea fi acceptabilă raţional la un moment *dar nu ca* adevărată.

Nu cred că raţionalitatea este definită prin intermediul unei mulţimi de 'canoane' sau 'principii' fixe; principiile metodologice au legătură cu concepţia noastră despre lume, inclusiv cu concepţia despre noi înşine ca parte a lumii şi se modifică în timp. Astfel, sunt de acord *cu filosofii subiectivişti că nu există un* organon *anistoric şi fix care defineşte ce înseamnă a fi raţional; dar, din faptul că perspectiva asupra raţiunii evoluează istoric, nu trag concluzia că raţiunea însăşi poate fi (sau poate evolua)* orice*, nici nu susţin un fel de compozit între relativism cultural şi 'structuralism' ca acela al filosofilor francezi. Dihotomia:* fie *canoane fixe anistorice ale raţionalităţii* fie *relativism cultural o consider depăşită.*

O altă trăsătură a acestei perspective este că raţionalitatea nu se restrânge la laboratoarele ştiinţifice şi nici nu diferă fundamental în laboratoarele ştiinţifice şi în afara lor. O asemenea concepţie este o derivaţie a pozitivismului, ce mizează pe ideea că lumea ştiinţifică este construită într-un mod din 'datele simţurilor' şi pe teza că termenii, în laboratoarele ştiinţifice, sunt 'definiţi operaţional'. Nu voi utiliza mult spaţiu pentru a critica perspectiva operaţionalistă şi pozitivistă asupra ştiinţei; acestea au fost atacate suficient în ultimii 20 de ani. Dar, ideea empiristă că 'datele simţurilor' constituie un fel de temei obiectiv pentru cel puţin o parte a cunoaşterii noastre va fi reexaminată în lumina a ceea ce vom spune despre adevăr şi raţionalitate (în Capitolul 3).

În linii mari, voi propune un punct de vedere în care mentalul nu este un simplu 'clişeu' al unei lumi care poate fi descrisă prin teoria adevărului unic. Pe de altă parte, perspectiva mea nu este nici aceea în care mentalul produce *lumea (sau o structurează după 'canoanele metodologice' şi 'datele simţurilor' independente de mental). Într-un limbaj metaforic, mentalul şi lumea împreună produc mentalul şi lumea. (Sau, pentru a face metafora mai hegeliană, Universul produce Universul – cu mental cu tot – împreună – jucând un rol aparte în propria sa producere).*

O ultimă trăsătură a abordării mele este următoarea: voi încerca să arăt că noţiunea noastră de raţionalitate este, la urma urmei, numai o parte a concepţiei despre umanitatea plenară, despre ideea noastră de bine. Adevărul este profund dependent de 'valori' (Capitolul 6). Iar ceea ce am spus despre raţionalitate şi istorie se aplică, la fel de bine, valorilor şi istoriei; nu există un dat anistoric, un set de 'principii morale' care definesc odată pentru totdeauna ce înseamnă umanitate plenară; dar nu rezultă că aceasta este în întregime culturală şi relativă. Starea actuală a teoriei adevărului – dihotomia între teoria adevărului corespondenţă şi abordările subiective – este, în viziunea mea, responsabilă cel puţin parţial pentru dihotomia între fapt şi valoare şi numai la un nivel foarte adânc şi corectând statutul adevărului şi raţionalităţii am putea să o depăşim. (O dihotomie care, aşa cum este înţeleasă de obicei, angajează la un anume relativism). Teoriile actuale ale adevărului sunt alienante; ele conduc la separarea sinelui de lume şi la imaginea unei lumi alcătuite din particule elementare aflate în vid (perspectiva 'fizicalistă', considerată ca descriere ştiinţifică *prin prisma teoriei adevărului unic) sau la o lume care constă din 'date ale simţurilor actuale şi posibile' (vechea perspectivă empiristă), sau la negarea existenţei vreunei lumi, opunându-i o sumă de viziuni pe care le creăm din motive variate (în principal neconştientizate). Scopul meu în această lucrare este să schiţez ideile directoare ale unei perspective non-alienante.*

Prelegerile Herbert Spencer, 'Filosofii şi înţelegerea umană' (susţinute la Unversitatea din Oxford în 1979) acoperă parţial textul de faţă, făcând parte dintr-un program mai vast, alături de studiul '„Dacă Dumnezeu este mort atunci totul este permis"... (reflexii asupra filosofiei limbajului)', Critique, *1980.*

Un grant al National Science Foundation[1] a suportat cercetarea pentru această carte pe parcursul anilor 1978–1980. Sunt recunoscător pentru acest ajutor.

---

[1] Pentru studiul 'The Appraisal of Scientific Theories: Historical versus Formal Mehodological Approaches'; *Agreement* Nr. SOC78–04276.

Thomas Kuhn şi Ruth Ana Putnam au studiat primele variante ale acestei cărţi, cărora le-au adus îmbunătăţiri, şi mi-au dat sfaturi înţelepte. De asemenea, am fost sprijinit de criticile şi sfaturile a numeroşi prieteni, între care Ned Block, David Helman şi Justin Leiber, ca şi de studenţi în cursul prelegerilor şi seminariilor de la Harvard. Câteva capitole au fost comunicate la Lima în vara lui 1980 (călătorie care a fost posibilă datorită unui grant obţinut de la Comisia Fulbright), iar Capitolul 2 a fost finalizat în cursul şederii la Lima. În această perioadă am beneficiat de discuţiile purtate cu Leopoldo Chiappo, Alberto Cordero Lecca, Henriques Fernandez, Francisco Miro Quesada şi Jorge Secada. Întreaga carte (într-o primă formă) a fost expusă în vara anului 1980 la Universitatea din Frankfurt şi ţin să mulţumesc colegilor mei de aici (în special lui Wilhelm Essler şi Rainer Trapp), ca şi grupului de studenţi şi celorlalţi prieteni din Germania (cu deosebire lui Dieter Henrich, Manon Fassbinder şi Wolfgang Stegmüller) pentru discuţiile încurajatoare şi stimulative.

Mulţumesc şi tuturor colegilor de la Departamentul de Filosofie de la Universitatea Harvard. În ultimii ani, am constatat o apropiere de puncte de vedere faţă de Nelson Goodman şi în perioada primelor variante ale cărţii am avut privilegiul să-i consult lucrarea Modalităţi de a construi lumea.

Sunt recunoscător şi faţă de Jeremy Mynott pentru încurajări şi sfaturi în calitate de editor.

Autorul

Cuprins

1 | Creiere în container 1
Teoriile magice ale referirii 4
Cazul creierelor în container 7
Testul Turing 11
Creiere în container (iarăşi) 16
Premisele argumentului 18
Raţiunile pentru a respinge conexiunea necesară dintre reprezentări şi referenţii lor 22

2 | O problemă privind referinţa 29
Intensiuni, extensiuni şi 'lumi noţionale' 34
Perspectiva tradiţională asupra interpretării 39
De ce nu funcţionează perspectiva tradiţională asupra interpretării 44
'Intrinsec' şi 'extrinsec' 48
'Supravieţuire' şi evoluţie 52
Intenţii: pur şi impur 56
Originea încurcăturii 59

3| Două doctrine filosofice 67
Internalism şi relativism 74
Teoria 'asemănării' 77
Turul de forţă a lui Berkeley 80
Viziunea lui Kant asupra cunoaşterii şi adevărului 81
Alternativa empiristă 87
Wittgenstein despre 'urmarea unei reguli' 90
Surprinderea 'formelor' şi asocierea empirică 93
Se poate greşi în ceea ce priveşte calitatea senzaţiilor trecute? 97
Din nou despre teoria adevărului corespondenţă 98

4| Minte şi corp 101
Paralelism, interacţionism, identitate 101
Teoria identităţii şi aprioricul 109
Creiere secţionate 114
Cât de bine este definită 'componenta calitativă'? 128
Realismul despre *qualia* 130

5| Două concepţii despre raţionalitate 135
Pozitivismul logic 138
Anarhism şi autorespingere 148
De ce relativismul este inconsistent 156
Ce ar fi de făcut? 162

6| Fapt şi valoare 167
Unele valori trebuie să fie obiective 176
Raţionalitatea în alte domenii 179
Super-benthamiştii 182
Subiectivismul despre bunătate 184
Autoritarism şi pluralism 191

7| Raţiune şi istorie 195

8| Impactul ştiinţei asupra concepţiei moderne privind raţionalitatea 227
Fetişismul 'metodei' 244

9| Valori, fapte şi cunoaştere 259
Teoria celor două componente 262
Moore şi 'eroarea naturalistă' 265
Din nou, 'nazismul raţional' 273

Anexă *281*

Referire şi relativism

Putnam Hilary Whitehall s-a născut la Chicago în 31 iulie 1926, tatăl său fiind Samuel Putnam, autor şi traducător. După 1934, familia sa se stabileşte în Philadelphia, iar în 1948 este licenţiat în matematică şi filosofie la Universitatea Pennsylvania. Devine doctor în filosofie la UCLA, sub îndrumarea lui H. Reichenbach, cu o disertaţie despre conceptul de probabilitate. La începutul carierei sale, Putnam predă matematica la Universitatea de Nord-Vest şi la Princeton, şi filosofia ştiinţei la MIT. În 1976 vine la Harvard, unde va rămâne până la retragere (2000), în cadrul departamentului de filosofie. A mai ţinut cursuri la Oxford, Londra şi Frankfurt.

Putnam Hilary este unul dintre cei mai importanţi filosofi ai secolului trecut, cu importante contribuţii în filosofia limbajului, filosofia minţii,

filosofia ştiinţei şi metafilosofie (în ultima perioadă a fost preocupat de locul şi importanţa socială a filosofiei). Iniţial se plasează pe poziţiile filosofiei analitice, suferind influenţe din partea lui H. Reichenbach şi R. Carnap. Ulterior, fără a părăsi stilul analitic, Putnam se manifestă ca neopragmatist sub influenţa lui Wittgenstein şi a lui Nelson Goodman. De fapt, Putnam şi-a schimbat opiniile filosofice de multe ori, asemănându-se, în această privinţă, cu B. Russell sau L. Wittgenstein. Unii comentatori l-au caracterizat pe Putnam (uneori pe nedrept), prin trei trăsături: a) are un stil ambiguu şi confuz, b) acordă prea mare atenţie unor exemple nerelevante, c) îşi schimbă foarte des opiniile.

Principalele sale lucrări sunt: *Mathematics, Matter and Method* (Cambridge, Marea Britanie, 1975), *Mind, Language, and Reality* (Cambridge, Marea Britanie, 1975), *Meaning and the Moral Sciences* (Routledge & Kegan Paul, Londra, 1978), *Reason, Truth and History* (Cambridge, Marea Britanie, 1981), *Realism and Reason* (Cambridge, Marea Britanie, 1983), *The Many Faces of Realism* (La Salle, Illinois, 1987), *Representation and Reality* (Cambridge, Massachussets, 1988), *Renewing Philosophy* (Cambridge, Massachussets, 1992), *The Threefold cord: mind, body and world* (Columbia, 1999) etc.

În filosofia limbajului, încă din deceniul VII al secolului al XX-lea, Putnam este adept al *externalismului semantic*. Această doctrină susţine că înţelesul termenilor constă în referinţa lor, respectiv, în relaţia cu realitatea externă. Alături de Saul Kripke, este unul dintre părinţii *teoriei cauzale* asupra referinţei, potrivit căreia o expresie dobândeşte referinţă printr-un eveniment singular, numit *botezul originar*, actul prin care expresia este introdusă în limbaj legată de referinţa sa. Ulterior, relaţia expresie–referinţă se conservă de la vorbitor la vorbitor, asemănător unui lanţ cauzal. Consecinţa acestei teorii este legătura constantă între expresii şi referinţă. Totuşi, o asemenea proprietate nu poate fi susţinută decât în cazul numelor proprii. Termenul 'om' se referă, în diferite contexte, la diferiţi oameni, aşa încât clasa sa nu este constantă.

În deceniul IX, Putnam depune eforturi pentru a respinge fizicalismul sau naturalismul în ce priveşte referinţa, care consideră că relaţia de referire este una fizică sau naturală. Nici aici nu are suficiente argumente; dacă relaţia dintre termen şi clasa sa nu este naturală, atunci de ce nu putem impune ca un termen să aibă clasa pe care o dorim noi? Nu cumva, clasa unui termen evoluează în mod natural? De pildă, putem împiedica în vreun fel ca mereu şi mereu alţi atomi de hidrogen să se combine cu alţi atomi de oxigen pentru a forma noi molecule de apă? Clasa termenului 'apă' evoluează natural, conform legilor fizico-chimice, aşa cum tot timpul unii oameni mor şi alţii se nasc, conducând la o evoluţie naturală a referinţei termenului 'om viu' etc. Lupta lui Putnam împotriva naturalismului în ce priveşte referinţa intră în conflict cu circumscrierea *termenilor naturali* care îi aparţine.

În filosofia minţii, Putnam susţine iniţial (deceniul VII) punctul de vedere că fenomenele mentale pot fi explicate prin analogia cu un computer, care, pe lângă partea materială (hardware) este caracterizat prin programele pe care le execută (software). Această perspectivă poartă numele de *funcţionalism*. Mergând cu consecvenţă, un timp, pe această cale, Putnam ajunge la absurdităţi, de pildă, că un spirit fără corp ar putea executa un anumit program, căzând în plin spiritism. Putnam însuşi ajunge cu timpul la concluzia că funcţionalismul nu dă seama de fenomenele mentale, explicând şi de ce.

Filosofia ştiinţei este abordată de Putnam îndeosebi din perspectiva disputei dintre realism şi empirism. Putnam respinge *realismul metafizic*, care este legat de teoria adevărului corespondenţă, argumentând că nu există o perspectivă absolută asupra realităţii şi adoptă *realismul intern*, potrivit căruia limbajul şi realitatea se află în interacţiune, aşa încât limbajul contribuie la constituirea realităţii, dar şi realitatea are rolul său în construirea limbajului. Adevărul apare, astfel, ca justificare epistemică idealizată. Dar teza că realitatea este modelată conceptual conduce la consecinţa că ar putea exista o realitate nemodelată conceptual (altfel, cum ne-am da seama, dacă nu prin comparaţie, că avem de-a

face cu o realitate modelată conceptual?). În acest caz, cum ar trebui să fie realitatea nemodelată conceptual, pentru a trage concluzia că realitatea este modelată conceptual? Putnam nu răspunde. Probabil, răspunsul cel mai potrivit este că nu există nici o realitate, nici modelată, nici nemodelată conceptual, dincolo de convenţiile sociale, aşa cum vom argumenta mai încolo.

Lucrarea de faţă (*Raţiune, adevăr şi istorie*), care este propusă cititorului român, înfăţişează stadiul ideilor filosofice ale lui Putnam la începutul deceniului IX asupra chestiunilor amintite: problema referinţei, problema raportului minte-corp, problema realismului şi relativismului, dar şi asupra altor teme precum raportul fapt-valoare, raţionalitatea şi filosofia istoriei etc. Ne oprim, în cele ce urmează, asupra tezelor susţinute de Putnam cu privire la referire şi relativism.

Referinţa este un parametru care caracterizează anumite expresii lingvistice. Putnam are tendinţa să considere că orice expresie, indiferent că este nume, termen sau propoziţie, are referinţă. El înţelege, prin referinţă, capacitatea unei expresii de a avea legătură cu ceea ce se află în exteriorul mentalului. O propoziţie este adevărată despre ceva în măsura în care se referă la acel ceva; un termen (sau un predicat) se aplică la un lucru, dacă se referă la acel lucru etc. Altă teză susţinută de Putnam este că nu expresiile ca atare au referinţă, ci acela care foloseşte o expresie 'se referă' la un lucru sau altul. Aici, de bună seamă, se simte influenţa lui Wittgenstein din *Investigaţii filosofice*. Aşa se explică argumentul că creierele din containere nu reuşesc să se refere la ele însele. De asemenea, schimbările în ce priveşte referirea sunt explicate prin aşa-numita *semantică a lumilor posibile*: cineva, printr-o expresie, se referă la diferite lucruri, în moduri diverse, în funcţie de lumea posibilă în care se află. Vom argumenta că aceste teze nu sunt necesare sau nu sunt suficiente pentru a da seama de fenomenele legate de referinţă.

Mai întâi, nu putem confunda toate funcţiile lingvistice pe care Putnam le trece sub titlul de referinţă. Modul în care o propoziţie are legătură cu lucrurile diferă de modul în care un nume sau un termen are legătură cu lucrurile; nu putem spune

că, de fiecare dată, avem de-a face cu referinţa. Trebuie să hotărâm care dintre aceste situaţii le numim referinţă, iar celelalte trebuie recunoscute sub alţi termeni.

De obicei, spunem că au referinţă propoziţiile. De pildă, propoziţia 'Aristotel este filosof' se *referă* la individul *denumit* 'Aristotel', iar propoziţia 'Toţi oamenii sunt muritori', se referă la toate fiinţele despre care spunem că sunt 'oameni', adică la elementele clasei oamenilor. În acest fel, convenim să numim *referinţă* un parametru care caracterizează propoziţiile, urmând ca parametrii celorlalte tipuri de expresii să-i distingem de referinţă.

Propoziţiile nu se caracterizează numai prin referinţă. Evoluţia valorii de adevăr a unei propoziţii nu poate fi explicată numai prin modificări ale referinţei. De pildă, în 1947 era adevărată propoziţia P = 'România este regat', care este falsă astăzi, în 2005, când este adevărată propoziţia Q = 'România este republică'. Referinţa celor două propoziţii este aceeaşi, atât în 1947 cât şi astăzi, ele se referă la acelaşi obiect, ţara numită 'România'. Totuşi, cele două propoziţii nu sunt echivalente, aşa încât trebuie să difere printr-un alt parametru, pe care îl numim *conţinutul* propoziţiei. P ropoziţiile P şi Q au aceeaşi referinţă dar un conţinut diferit. La o analiză sumară, se constată că ele au acelaşi subiect, dar au predicate diferite, aşa încât este firesc să presupunem că subiectul este cel care determină referinţa unei propoziţii, pe când predicatul stabileşte conţinutul. Prin urmare, referinţa depinde de funcţiile lingvistice îndeplinite de expresiile care joacă rolul de subiect.

Referinţa unei propoziţii care are ca subiect un nume propriu se comportă altfel decât referinţa propoziţiilor care au ca subiect termeni. În vreme ce o propoziţie precum 'Aristotel este filosof' are întotdeauna aceeaşi referinţă, propoziţia 'Toţi oamenii sunt albi' se referă mereu la alte lucruri, deoarece, odată cu trecerea timpului, oamenii sunt alţii. Din acest motiv, trebuie să distingem între modul în care se raportează la lucruri numele şi termenii. Despre nume spunem că îi corespunde un *denotat*, iar despre termeni spunem că le corespunde o *clasă*.

Dacă subiectul unei propoziţii este un nume, atunci referinţa ei coincide cu denotatul numelui. Un nume are, în orice context, acelaşi denotat. Aceasta reiese din faptul că propoziţiile care au ca subiect nume nu îşi modifică referinţa la negaţie. Bunăoară, propoziţiile 'Aristotel este filosof' şi 'Aristotel nu este filosof' au aceeaşi referinţă. Dacă nu ar fi aşa, cele două propoziţii nu s-ar afla în relaţie de contradicţie. Potrivit principiului terţului exclus, în orice context, una dintre aceste propoziţii trebuie să aibă loc, ceea ce înseamnă că, în orice context, numele 'Aristotel' are acelaşi denotat. De aici rezultă că denotatul numelui Aristotel, adică obiectul numit astfel, există perpetuu. Faptul că un nume are acelaşi denotat indiferent de împrejurări, a fost observat de Saul Kripke, care vorbeşte de *designare rigidă*. Din acest motiv, propoziţiile care presupun sinonimia numelor proprii, dacă sunt adevărate, atunci sunt adevărate în orice context, adică sunt *necesare* (această proprietate a numelor proprii nu poate fi extinsă asupra termenilor, aşa cum încearcă Putnam în exemplul cu 'temperatura'). Pe de altă parte, coincidenţa denotatelor este stabilită empiric, aşa încât există propoziţii necesare empirice. Astfel, denotatul este, pentru nume, un parametru constant, de aceea funcţia lingvistică a numelui este de a fixa un obiect, de a fixa denotatul său. Înţelesul unui nume constă în denotatul său.

Pentru a explica de ce o propoziţie precum 'România este regat' îşi schimbă valoarea de adevăr (era adevărată în 1947, dar este falsă acum, în 2005), trebuie să recunoaştem că, pe lângă denotat, care a rămas acelaşi, numele se caracterizează printr-un al doilea parametru, pe care îl numim *sens* şi care, spre deosebire de denotat, se schimbă în timp, adică este un parametru *dinamic*. Numele nu are rolul de a fixa sensuri, aşa încât, în vreme ce denotatul este un parametru *intenţional*, sensul unui nume este un parametru *natural*, adică evoluează în afara voinţei noastre.

Forma generală a unei propoziţii care are ca subiect un termen este: 'kS sunt αP', unde k este un număr cardinal, numit *cuantificator*, iar α este un număr raţional (sau o expresie compusă, pornind de la un număr raţional, cum ar fi 'π', '$\sqrt{2}$' etc.) care are

rolul de a fixa locul termenului predicativ într-o scală. De pildă, în propoziţia 'Zece oameni din această sală au 1,80 m înălţime', 'Zece' este un număr cardinal care are rolul de cuantificator, în vreme ce '1,80' indică un termen pe scala înălţimilor. Cuantificatorii universali sau particulari pot fi reduşi la numere cardinale. Cuantificatorul universal negativ 'nici un' este, de fapt, numărul cardinal 'zero', pe când cuantificatorul universal afirmativ 'toţi', 'orice' etc. este cardinalul clasei corespunzătoare termenului subiect: 'card *S*'. Bunăoară, 'Toţi *S* sunt αP' este acelaşi lucru cu 'Card(*S*) *S* sunt αP'. Cuantificatorului particular afirmativ 'unii', îi corespunde intervalul închis la dreapta de numere cardinale (0, card(*S*)], pe când celui particular negativ îi corespunde intervalul închis la stânga [0, card(*S*)).

Referinţa unei propoziţii care are ca subiect un termen nu coincide întotdeauna cu clasa acelui termen. De pildă, propoziţia 'Toţi oamenii sunt albi' se referă la clasa oamenilor, pe când 'Unii oameni sunt albi' se referă la o parte nevidă a clasei oamenilor. Referinţa acestor două propoziţii trebuie să fie diferită, deoarece pot fi situaţii în care valoarea lor de adevăr să difere deşi conţinutul lor este acelaşi.

Pe de altă parte, referinţa propoziţiilor care au ca subiect termeni se schimbă în timp. De pildă, propoziţia 'Toţi oamenii sunt albi' are în fiecare moment altă referinţă deoarece clasa oamenilor este diferită, la moment diferite. De fapt, la fiecare moment, avem de-a face cu o altă clasă a oamenilor, după cum reprezentanţii săi se nasc sau se descompun, respectiv, intră sau ies din clasa oamenilor. Teza, pe care o susţine şi Putnam, că clasa corespunzătoare unui termen, precum 'om', conţine în fiecare moment toţi oamenii, inclusiv cei care au existat sau vor exista vreodată, nu poate fi susţinută. În acest caz, o clasă ar avea elemente care nu există. O propoziţie precum 'Nu există dinozauri' nu ar putea fi adevărată în nici un context, deoarece s-ar referi inclusiv la dinozaurii din Jurasic. Astfel, trebuie acceptat că referinţa unei propoziţii care are ca subiect un termen se

modifică în timp. Bunăoară, o propoziţie precum 'Filosoful născut la Stagira a scris *Organonul*' nu are referinţă înainte de naşterea lui Aristotel sau chiar înainte ca acesta să devină filosof, dar are referinţă după aceea.

Din aceste exemple rezultă că termenilor nu le corespunde o clasă fixă, adică termenii nu au funcţia lingvistică de a fixa o clasă; ci, printr-un termen se fixează un sens. Sensul este, în ce priveşte termenii, parametrul constant, sau intenţional, în vreme ce clasa este parametrul dinamic. Înţelesul unui termen constă în sensul său. Faptul că termenii nu fixează obiecte, ci sensuri, este întărit de propoziţiile care au valoare de adevăr fără a avea referinţă. De exemplu, despre propoziţia 'Pegas este un cal înaripat' nu s-ar putea spune că este falsă, deşi nu există cai înaripaţi, deoarece prin 'Pegas' tocmai asta se înţelege – un cal înaripat. La fel, propoziţia 'Zeus este şeful zeilor din Olimp' trebuie considerată adevărată, dar de aici nu rezultă că există vreun astfel de zeu, ci numai că sensul termenului 'Zeus' are drept componentă 'şeful zeilor din Olimp', pe care o păstrează în împrejurările prevăzute de mitologia vechilor greci. Prin 'Zeus este un personaj mitologic' trebuie să înţelegem tocmai că 'Zeus' este un termen, nu un nume, al cărui sens este precizat prin intermediul unui text care serveşte ca definiţie.

Vedem, astfel, că între propoziţiile care au referinţa determinată printr-un nume şi cele care au ca subiect un termen, există diferenţe esenţiale, aşa încât funcţiile lingvistice ale propoziţiei, numelor sau termenilor nu pot fi confundate. Dacă ne hotărâm să numim 'referinţă' componenta propoziţională respectivă, trebuie să distingem prin alţi termeni rolurile avute de nume şi termeni; spunem că numele au *denotat*, iar termenilor le corespund *clase*, fără a le confunda între ele.

O altă confuzie prezentă la Putnam, sub influenţa lui Wittgenstein, este cea dintre *referinţă* şi *referire*, autorul american având în vedere cu deosebire referirea. Dacă referinţa este un parametru care caracterizează propoziţiile, referirea poate fi înţeleasă ca o *acţiune* întreprinsă de locutor, atunci când apelează la

o propoziţie. De pildă, cineva care rosteşte 'România este republică', *se referă* la România. Faptul că referinţa nu poate fi confundată cu referirea reiese din următorul exemplu. Să presupunem că doi subiecţi care rostesc propoziţiile:

(1) România este republică.

(2) Ţara care are capitala la Budapesta este republică.

doresc să se refere la aceeaşi ţară, România. În vreme ce prima propoziţie are referinţa România, a doua propoziţie nu are aceeaşi referinţă, ci se referă la Ungaria. Situaţia descrisă nu este 'ştiinţifico-fantastică', ci se poate uşor petrece dacă cineva nu ştie care este capitala României, considerând că este Budapesta. Al doilea vorbitor *intenţionează* să se refere la România, dar nu reuşeşte. Observăm că uneori, referirea nu coincide cu referinţa, diferenţele putând fi explicate fie prin necunoaştere, fie printr-o anumită intenţie. De pildă, al doilea vorbitor s-ar putea să ştie foarte bine că Budapesta este capitala Ungariei, dar să facă o glumă, referindu-se astfel la România. Altă situaţie este aceea în care cineva utilizează metafore pentru a se referi sau când referirea rămâne *ocultă*, ascunsă pentru majoritatea ascultătorilor, cu excepţia câtorva care împărtăşesc cu emitentul un anumit *cod*, adică anumite convenţii opace pentru alţii.

Observăm că *referirea* diferă de multe ori de referinţă; mai mult, ca orice acţiune, poate avea succes sau nu, adică, destinatarul s-ar putea să descifreze corect referirea sau eronat. Astfel, în vreme ce referinţa este accesibilă prin *interpretare*, referirea nu este accesibilă în nici un fel altui subiect decât vorbitorului însuşi. Chiar dacă l-am întreba 'la ce te-ai referit în propoziţia (2)?', subiectul ar putea minţi şi spune 'la Ungaria', deşi el intenţiona să se refere la România. Nici un test (nici măcar *testul Turing pentru referire*) nu va putea descifra adevărata referire.

Faţă de emitent, discursul primeşte autonomie, aşa încât, poate fi interpretat fără a şti care au fost intenţiile emitentului. Dacă ar trebui de fiecare dată să cercetăm aceste intenţii, atunci

înţelegerea ar fi imposibilă. Interpretarea presupune anumite constrângeri, aşa cum constată şi Putnam, putând fi *literală* sau *metaforică*. În primul caz, pentru a interpreta, sunt folosite *convenţiile sociale* obişnuite care planează asupra expresiilor lingvistice, în vreme ce în al doilea, se ia în considerare *libertatea* emitentului de a modifica aceste convenţii. Odată ce, aşa cum am văzut, denotatul numelor şi sensul termenilor sunt parametri intenţionali, înseamnă că oricine poate înzestra orice nume cu orice denotat şi orice termen cu orice sens, aşa încât este posibil ca aceeaşi expresie să primească interpretări diferite în texte diferite.

Astfel, Putnam nu va şti niciodată dacă furnica, prin urma lăsată pe nisip: 'Winston Churchill a fost premierul Marii Britanii în al doilea Război Mondial' s-a referit sau nu la Winston Churchill. Argumentul că furnicile nu sunt în stare să se refere nu poate fi adus fără a cădea sub acuzaţia de 'şovinism' pe care Putnam caută să o evite cu alte ocazii. Dacă o furnică este în stare să traseze pe nisip o asemenea urmă, aşa cum presupune Putnam, de ce să nu o credem în stare de referire? La fel, argumentul că furnica nu se poate referi la Churchill pentru că nu l-a văzut niciodată sau nu a auzit niciodată de el nu are susţinere. Dacă nu am văzut niciodată balene nu înseamnă că nu mă pot referi la ele. Mendeleev s-a referit cu succes la elemente chimice pe care nu le-a văzut şi despre care nu a auzit vorbindu-se niciodată. Discuţia rămâne, însă, la nivel de presupunere, pentru că, repet, nimeni nu poate determina intenţionalitatea *referirii* cuiva. În schimb, Putnam poate analiza dacă urma pe nisip despre care vorbim are *referinţa* Winston Churchill, independent de referire sau de succesul referirii (poate furnica dorea să se refere la A. N. Chamberlain). Dacă suntem adepţii unei interpretări literale, atunci urma lăsată de furnică pe nisip are referinţa Winston Churchill; dacă admitem varianta metaforică, s-ar putea ajunge la concluzia, în absenţa altor elemente, că propoziţia are ca referinţă, de fapt, o altă furnică, dintr-un alt muşuroi. În concluzie, propoziţia trasată de furnică poate avea referinţa Winston Churchill în anumite interpretări; dacă furnica a intenţionat sau nu să se refere la Winston Churchill prin urma

respectivă, nu putem şti, aşa cum nu putem şti nici despre oameni dacă intenţionează să se refere la ceva sau avem de-a face cu simple coincidenţe.

Pentru a explica schimbarea valorii de adevăr a propoziţiilor (adică a referinţei sau predicaţiei lor) Putnam apelează la *semantica lumilor posibile*. Potrivit ipotezei lumilor posibile, pe lângă lumea reală, pot exista şi alte lumi faţă de care o propoziţie adevărată în lumea reală poate fi falsă sau o propoziţie falsă în lumea reală poate fi adevărată. Astfel, dacă p este o propoziţie adevărată în lumea reală, poate exista o altă lume în care $\sim p$ (negaţia propoziţiei p) nu este falsă. De exemplu, dacă în lumea reală are loc 'Aristotel este filosof', într-o lume posibilă poate avea loc 'Aristotel nu este filosof'.

Pentru a da o soluţie problemelor de filosofia limbajului sau chiar unor probleme aparţinând altor domenii ale filosofiei, Putnam apelează la ipoteza lumilor posibile, care devine, astfel, un fel de panaceu universal, jucând un rol analog cu *deus ex machina* din vechile piese de teatru. Fără să dea vreo demonstraţie de existenţă pentru lumile posibile, Putnam le invocă de câte ori are prilejul, ba mai mult, diversitatea lor este covârşitoare. Astfel, pe lângă lumi posibile logic, pentru Putnam există lumi posibile filosofic, lumi posibile matematic sau lumi posibile fizic, fără ca acestea să se confunde în vreun fel şi fără a fi lămuriţi asupra relaţiilor dintre ele. Care sunt, însă, criteriile după care sunt deosebite şi câte asemenea tipuri de lumi posibile pot fi, autorul nu ne spune. Evidenţierea atâtor feluri de lumi posibile conduce la teza adevărului multiplu: o propoziţie poate fi adevărată logic, dar falsă fizic, poate fi adevărată fizic dar falsă filosofic etc., doctrină care se apropie până la identificare cu aceea a dublului adevăr susţinută de unii filosofi care nu voiau să intre în conflict cu ideologia dominantă la un moment dat, aşa încât multiplicarea tipurilor de lumi posibile, care este totuna cu multiplicarea adevărurilor, este o doctrină ideologică, nicidecum filosofică.

Chiar dacă ne-am restrânge la un singur fel de lumi posibile, conceptul de lume posibilă se dovedeşte *inconsistent*, ori acesta este un motiv serios de respingere, pe care Putnam îl

acceptă. Nici un partizan al semanticii lumilor posibile nu a formulat vreo teoremă de existenţă şi nici nu va formula, pentru că se poate dovedi că lumile posibile, dimpotrivă, nu există.

Criteriile după care pot fi deosebite lumile posibile sunt două: a) odată cu trecerea de la o lume la alta se modifică referinţa propoziţiilor nominale (care au ca subiect nume, adică propoziţiile de forma 'Fa'); b) odată cu trecerea de la o lume la alta se schimbă conţinutul propoziţiilor nominale.

Potrivit primului criteriu, atunci când trecem de la lumea *L*1 la lumea posibilă *L*2, referinţa unei propoziţii 'Fa' devine alta, respectiv, numele *a* are un denotat în lumea *L*1 şi alt denotat în lumea *L*2. Aşa se petrec lucrurile în exemplul celebru a lui Putnam, cu 'apa' şi cu Pământul geamăn. Deşi expresia 'apă' nu este un nume, pentru Putnam, acesta este un termen natural care se comportă ca un nume. Pământul geamăn poate fi asimilat unei alte lumi posibile. Fie propoziţia 'Apa este limpede', adevărată în ambele lumi posibile.

Dacă *L*1 este lumea noastră, referinţa propoziţiei este lichidul care are compoziţia H_2O, în vreme ce în *L*2 (lume posibilă), lichidul la care se referă este altul, având compoziţia *XYZ*. Dar acesta nu este un criteriu pentru a deosebi între lumi diferite, deoarece cele două propoziţii pot avea loc în aceeaşi lume, pentru că nu sunt în opoziţie. Deoarece termenii sunt arbitrari (cu unele excepţii, cum ar fi chiar termenul H_2O), nimic nu împiedică doi subiecţi să folosească termenul 'apă' cu privire la două lichide diferite în aceeaşi lume. Diferenţa de referinţă nu este suficientă pentru a presupune existenţa unor lumi diferite. În ciuda coincidenţei privind sunetele folosite, nu avem de-a face cu una şi aceeaşi propoziţie, deoarece diferă prin parametrul referinţei. Putnam se foloseşte de faptul că, pe câtă vreme 'apă' este arbitrar, (se poate constata prin aceea că în limbi diferite din *lumea reală* acelaşi lichid este cunoscut sub termeni diferiţi ('apă', 'water' etc.)), termenul 'H_2O', nu este cu totul arbitrar. Odată ce s-au stabilit termeni pentru hidrogen şi oxigen, ca şi o regulă de obţinere a termenilor pentru substanţele compuse, apa nu poate primi alt termen decât 'H_2O'. Dar, deoarece tabelul lui

Mendeleev este complet, chiar pe planeta presupusă de Putnam (o dată ce este similară Pământului), ar trebui să se găsească substanţa H_2O, altfel, fie planeta nu ar fi similară Pământului, cum cere ipoteza, fie pe acea planetă, tabelul lui Mendeleev ar fi altul, dar Putnam nu ne spune care este. În acest fel, oricine, prin *analiză chimică* – aceasta nu este decât o formă de analiză semantică – poate trage concluzia că lichidul numit 'apă' pe Pământul geamăn este altul decât lichidul pe care îl numim 'apă' pe Pământ, iar acesta fie nu ar avea nume (nefiind descoperit), fie ar fi numit altfel. Un bun dicţionar care ar permite traducerea dialectului de pe planeta geamănă în dialectul nostru, ar soluţiona problema. Astfel, dacă subiectul este interpretat în aşa fel încât propoziţia să aibă o altă referinţă, ipoteza lumilor posibile este inutilă, deoarece noua interpretare nu intră în opoziţie cu cea veche, deoarece, pentru ca două propoziţii nominale să fie incompatibile, trebuie să aibă aceeaşi referinţă. Nu este nevoie să se presupună un alt Pământ pentru a întâlni fenomene precum cele invocate de Putnam. Chiar pe Pământul nostru, în lumea reală, există situaţii în care acelaşi cuvânt este folosit pentru a ne referi la lucruri diferite. Propoziţia 'Broasca este un mecanism' este adevărată într-un manual de lăcătuşărie şi falsă într-un manual de zoologie relativ la *lumea reală*, dar referinţa este diferită. Lumile posibile nu pot fi deosebite după diferenţa de referinţă pentru că în una şi aceeaşi lume, care este lumea reală, se constată acest fenomen.

Al doilea criteriu după care pot fi deosebite lumile posibile este diferenţa de conţinut. Astfel, două propoziţii care au aceeaşi referinţă, dar au conţinut opus, nu pot fi adevărate împreună. Potrivit principiilor logicii, 'Fa' şi '~Fa' nu *pot* fi adevărate în acelaşi context. Totuşi, dacă 'Fa' este adevărată *factual*, adevărul negaţiei sale nu este exclus; de aici, ipoteza că negaţia este *posibil* adevărată, adică există o *lume*, alta decât lumea reală, dar care este posibilă, unde această propoziţie este adevărată. (Termenul 'lume' este folosit pentru a explica de ce o propoziţie este adevărată).

Primul criteriu prin care sunt deosebite lumile posibile poate fi redus la al doilea. Faptul că o propoziţie precum 'Fa' are referinţă diferită în lumi diferite este acelaşi lucru cu a spune că într-o lume este adevărat că 'Propoziţia "Fa" se referă la *x*', iar în altă lume, este adevărat că 'Propoziţia "Fa" nu se referă la *x*', aşa încât, obiecţiile privind al doilea criteriu resping şi primul criteriu prin care pot fi deosebite lumile posibile.

Dar, pentru ca o lume să fie posibilă, ea trebuie să *poată avea loc*. Dacă o lume nu poate avea loc, de bună seamă că este anormal să o considerăm posibilă, cel mult, îi putem spune *imposibilă*.

Fie *L*1, lumea *reală* la momentul t_1, în care are loc 'Fa' şi fie *L*2 lumea posibilă în care s-ar petrece '~Fa'. Pentru ca *L*2 să fie posibilă, este nevoie ca ea să poată avea loc la un moment t_2. Aceste două lumi nu pot avea loc simultan, (când $t_2 = t_1$), deoarece atât 'Fa' cât şi '~Fa' ar fi adevărate în acelaşi context, ceea ce contravine principiului noncontradicţiei. Dacă cele două propoziţii ar avea loc la momente diferite (adică în contexte diferite), ele nu ar fi incompatibile, respectiv, nu ar fi nevoie să se presupună mai multe lumi posibile, ci, relativ la aceeaşi lume, o propoziţie ar fi adevărată într-un context, iar cealaltă în alt context; lucrurile s-ar rezolva spunând că *starea obiectului la care se referă propoziţia* (*a*), a suferit o *schimbare*. Prin urmare, apelând la conceptele de 'stare' şi 'schimbare', caracterul factual al propoziţiilor poate fi explicat în cadrul unei singure lumi, fără a fi nevoie să se apeleze la aşa-numitele 'lumi posibile'. În ultimă instanţă, o singură lume, cea reală, este suficientă pentru a da seama atât de schimbările de referinţă cât şi de cele de conţinut ale propoziţiilor.

De exemplu, dacă în lumea reală este adevărată propoziţia 'Napoleon a pierdut bătălia de la Waterloo', ne putem întreba, când era posibil ca Napoleon să fi câştigat această bătălie (când putea avea loc lumea posibilă în care Napoleon să fie victorios)? De bună seamă, o asemenea eventualitate nu se putea petrece înainte de încheierea bătăliei – ar însemna că Napoleon era capabil să câştige o bătălie chiar înainte de a avea loc, atribuin-

dă-i-se marelui general, capacităţi supranaturale. Nici după bătălie Napoleon nu o putea câştiga – ar însemna că atât a câştigat bătălia, cât şi a pierdut-o, ceea ce ar contraveni principiului noncontradicţiei. Astfel, odată ce *în realitate*, Napoleon a pierdut bătălia de la Waterloo, niciodată, nici înainte, nici după bătălie, nu ar fi putut să o câştige, adică, lumea în care Napoleon a câştigat bătălia de la Waterloo nu poate avea loc, respectiv, este o lume *imposibilă*, aşa încât a susţine că într-o asemenea lume acea propoziţie ar fi adevărată ar însemna că imposibilul poate fi adevărat!

În consecinţă, nu poate exista decât o singură lume, cea reală. Dictonul lui Leibniz trebuie modificat: trăim în singura lume posibilă şi nu în cea mai bună dintre lumile posibile. O dată ce un fapt s-a petrecut, este *imposibil* să se fi petrecut în alt fel sau să nu se fi petrecut. Aşa-zisele lumi posibile se dovedesc a fi imposibile sau *contrafactuale*. Le putem numi *lumi imaginare*. Astfel, ne putem *imagina* că Napoleon a câştigat bătălia de la Waterloo şi ne putem întreba 'ce s-ar fi întâmplat dacă...', dar suntem conştienţi că o asemenea eventualitate este *imposibilă*. Imaginarul se deosebeşte de posibil. Putem să ne imaginăm cai zburători sau monştri marini cu 99 de capete sau centauri (îi putem chiar desena), dar asta nu înseamnă că aşa ceva poate să existe. Imaginarul este materia basmului; multe dintre exemplele lui Putnam pe care le plasează în lumi *posibile* au loc, de fapt, în asemenea lumi imaginare, care nu au relevanţă pentru filosofie. Aşa este lumea în care nu există alte plante decât muşchi, dar există baldachine şi clopotniţe – oare din ce or fi făcute? – sau lumea în care milioane de ani o maimuţă acţionează tastele unei maşini de scris – în acest răstimp, ori moare maimuţa, ori evoluează în filosof. Desigur, imaginând tot felul de asemenea lumi, se pot susţine orice teze dorim. Dificultatea este că, lumile nefiind *posibile*, nici tezele respective nu găsesc vreun temei.

Putnam se ridică viguros împotriva relativismului. El înţelege relativismul, urmându-l pe Protagoras, drept teza că, pentru o propoziţie *p*, din propoziţia 'Eu cred *p*' se poate infera *p* sau că 'Eu cred *p*' este o *condiţie suficientă* pentru *p*, aşa încât este

suficient să cred p pentru ca credinţa în p să fie întemeiată sau, altfel spus, 'Orice credinţă este întemeiată'. A spune că o propoziţie este relativă faţă de *mine*, ar însemna că, dacă *eu* cred acea propoziţie, ea este adevărată. Adică,

$$\text{'Relativ}_x(p) \equiv (\text{Cred}_x(p) \vdash p)\text{'}$$

(p este relativă faţă de x dacă şi numai dacă, atunci când x crede p, are loc p). Faţă de această teză, Putnam aduce următoarele contraargumente:

1. Teza relativistă se autorespinge. Dacă relativistul este consecvent, trebuie să accepte că inclusiv propoziţia 'Toate propoziţiile sunt relative' este relativă, că punctul său de vedere (care este perspectiva relativistă) nu este mai bun decât punctul de vedere opus.

2. Obiecţia lui Platon. Platon îl acuză pe Protagoras de regresie la infinit. Propoziţia p poate fi chiar 'Eu cred p', când nu se poate determina condiţia suficientă pentru p, ajungându-se la regresia 'Eu cred că eu cred că...eu cred că...'.

3. Relativismul este inconsistent, permiţând justificarea unor propoziţii false. Dacă toţi oamenii ar avea aceleaşi credinţe, nu ar exista nici o diferenţă între relativişti şi non-relativişti. Pentru ca relativismul să aibă sens, trebuie să existe diferenţe de opinii. În acest caz, un subiect X, poate să creadă p, în vreme ce un alt subiect, Y, poate să creadă $\sim p$. Din teza relativistă, urmează:

$$\begin{array}{ll} X \text{ crede } p. & Y \text{ crede } \sim p. \\ **p & **\sim p \end{array}$$

potrivit principiilor logicii, una dintre concluzii trebuie să fie falsă, aşa încât, teza relativistă apare ca inconsistentă.

4. Relativismul conduce la *solipsism epistemologic*, adică la teza că numai eu am dreptate, iar toţi ceilalţi se înşeală. Acest argument este inspirat de Wittgenstein. Dacă credinţele mele sunt întemeiate, atunci, una din două: credinţele celorlalţi coincid cu ale mele sau sunt diferite de credinţele mele. În primul

caz, nu ar fi vorba de relativism, ci de credinţe împărtăşite de toţi; în al doilea caz, odată ce credinţele mele sunt întemeiate, credinţele celorlalţi sunt eronate. Urmează ca, dacă acceptăm relativismul, ceilalţi nu pot avea credinţe întemeiate. Astfel, numai eu sunt capabil de cunoaştere şi comunicare, în vreme ce semenii mei nu sunt în stare decât să rostească zgomote fără înţeles sau fără conţinut cognitiv.

La prima obiecţie se poate contraargumenta că a fi relativ, nu înseamnă a fi fals. O propoziţie poate fi relativă fără a fi falsă în raport cu vreun evaluator. Chiar dacă este relativă faţă de cei care o evaluează, s-ar putea ca propoziţia în cauză să fie adevărată faţă de toţi. Asemenea propoziţii, care sunt adevărate pentru oricine şi în orice moment, există. De pildă, fie propoziţia 'Bucureşti este la Est', care este adevărată pentru un evaluator din Timişoara, dar este falsă pentru unul aflat la Constanţa. Dacă precizez starea evaluatorului, se ajunge la 'Pentru oricine aflat în Timişoara, Bucureşti este la Est', care este adevărată pentru toţi evaluatorii, indiferent unde s-ar afla. O propoziţie este relativă faţă de un factor x atunci când valoarea ei de adevăr depinde de x. De pildă, propoziţiile pot fi relative faţă de timp; 'România este regat' era adevărată în 1947, dar este falsă astăzi. În acest caz, timpul trebuie considerat ca o variabilă, alături de variabilele individuale şi cele predicative. Dacă această variabilă este cuantificată sau i se dă o valoare, se obţin propoziţii care au aceeaşi valoare de adevăr în raport cu orice moment temporal, cum ar fi propoziţia 'În 1947, România era regat'. Între factorii de care depinde valoarea unei propoziţii se numără şi evaluatorul. Astfel, dacă propoziţia 'Toate propoziţiile sunt relative' este relativă în raport cu cei care o evaluează, se poate construi propoziţia 'Pentru Protagoras, toate propoziţiile sunt relative' care este adevărată (falsă) faţă de orice evaluator. O asemenea propoziţie este relativă faţă de evaluatori, dar are aceeaşi valoare de adevăr în raport cu oricare dintre aceştia. Situaţia este asemănătoare *funcţiilor constante* din matematică.

Cu privire la obiecţia lui Platon, Putnam observă corect că acesta este un argument slab, care nu afectează relativismul. Deşi este *posibil* ca şirul operatorilor 'cred că' să continue la nesfârşit, nu este obligatoriu. Protagoras se poate opri la ce nivel al credinţei doreşte pentru a infera *p*. Pe de altă parte, abia dacă nu s-ar accepta că măcar pentru unele propoziţii are loc teza lui Protagoras, s-ar ajunge la regresia amintită. Pentru a nu cădea în regresie, este nevoie ca, măcar pentru propoziţiile de credinţă, Protagoras să aibă dreptate. Dacă din 'Cred că cred că *p*' derivă 'Cred că *p*', atunci regresia se stinge, orice stare de credinţă descrisă prin mai mulţi operatori 'cred' s-ar reduce la starea descrisă prin unul singur. În alt caz, starea de credinţă descrisă de o infinitate de asemenea operatori ar fi ireductibilă. În plus, chiar în interiorul gnoseologiei lui Platon, teza lui Protagoras este prezentă implicit. Pentru Platon, a cunoaşte că *p* înseamnă a crede întemeiat *p* la care trebuie adăugat adevărul lui *p*. Pentru a nu cădea în regresie la infinit, este nevoie ca, măcar pentru unele propoziţii să aibă loc 'Cred $p \vdash p$'. În acest fel, *p* este cunoscut atunci când este crezut, deoarece au loc atât credinţa întemeiată în *p*, cât şi adevărul lui *p*. Dacă nu s-ar întâmpla astfel, atunci ce cunoştinţe am putea avea, potrivit teoriei lui Platon?

Pentru a respinge obiecţia că relativismul este inconsistent, este nevoie de o analiză mai amănunţită. Mai întâi, pentru a nu cădea în scepticism, trebuie acceptat că unele propoziţii *trebuie* să fie relative. Pe parcursul lucrării sale, Putnam lasă să se întrevadă numeroase dovezi că nu este sceptic. Scepticul afirmă că nici una dintre credinţele noastre nu este întemeiată. O credinţă într-o propoziţie *p* este întemeiată dacă acea propoziţie este adevărată şi dacă putem stabili că este adevărată. Scepticul spune că nu avem nici un mijloc de a trece dincolo de credinţe, adică nu suntem îndreptăţiţi la nici o inferenţă de tipul (IC): 'Eu cred $p \vdash p$'. Astfel, pentru a combate scepticismul, trebuie să arătăm că, măcar pentru unele propoziţii, o asemenea inferenţă este corectă.

Dacă nici una dintre credinţele noastre nu ar fi întemeiată, nu s-ar explica de ce avem succes în unele dintre acţiunile

noastre; în termenii lui Putnam, nu s-ar explica de ce supraviețuim. Scepticul scapă de această obiecție arătând că atitudinea rațională ar fi abținerea de la orice acțiune. Totuși, chiar și abținerea de la acțiune este o acțiune, așa încât ar trebui să se explice de ce o asemenea strategie are succes, respectiv, de ce credința că nu există credințe întemeiate este întemeiată? Prin urmare, respingând relativismul, Putnam subminează teza că unele dintre credințele noastre trebuie să fie întemeiate pentru a explica supraviețuirea, pe care o susține în altă parte.

Inferențele de forma IC pot fi corecte numai pentru propozițiile p pentru care nu are loc obiecția lui Putnam, adică pentru care nu se ajunge la inconsistență. Am văzut că obiecția de inconsistență adusă relativismului constă în aceea că se pot construi raționamente de forma IC care conduc la concluzii false. Dacă explicităm variabila propozițională p printr-o schemă de propoziții simple, Fz, atunci raționamentele imaginate de Putnam sunt:

X crede Fz.	Y crede $\sim Fz$.
$^{**}Fz$	$^{**}\sim Fz$

Există, totuși, o situație în care asemenea raționamente nu pot genera concluzii contradictorii, în cazul în care $z = x$ (argumentul propoziției crezute coincide cu argumentul credinței), când se obțin raționamente de forma (ICE): 'X crede $Fx \vdash Fx$'. De data aceasta, inconsistența invocată de Putnam nu mai apare. Fie perechea de raționamente:

X crede Fx.	X crede $\sim Fx$.
$^{**}Fx$	$^{**}\sim Fx$

Premisele celor două raționamente nu pot fi adevărate împreună în același timp, deoarece s-ar încălca principiile logicii: unul și același individ s-ar afla în două stări incompatibile simultan sau, la același moment, s-ar afla și nu s-ar afla într-o anume stare (ar crede și nu ar crede o propoziție). În această situație, tipul de raționament de mai sus poate fi corect fără a conduce la inconsistență, deoarece, pentru a detașa concluzia, nu este suficientă corectitudinea raționamentului, ci trebuie adăugat

adevărul premisei, or premisele celor două raţionamente nu pot fi adevărate simultan, deci nici concluziile nu pot fi detaşate simultan, aşa încât inconsistenţa nu este necesară. Ea poate apare numai dacă, atunci când premisa este adevărată, concluzia este falsă, respectiv, numai dacă raţionamentul nu este corect.

Să arătăm că raţionamentele de forma ICE sunt corecte. Premisa raţionamentului are forma '*X* crede că *x* simte *F*', de pildă, o asemenea premisă este: 'Eu cred că eu simt roşu', adică presupune o credinţă despre stările proprii sau interne. Stările proprii sunt accesibile şi, în plus, sunt singurele accesibile subiectului. Acestea sunt stările *conştientizate* de către subiect; a conştientiza o stare proprie şi a te afla într-o stare proprie este acelaşi lucru. Conştientizarea are loc în prezent, deoarece o stare trecută nu poate fi conştientizată pentru că nu mai există, iar una viitoare nu poate fi conştientizată pentru că încă nu există. Dacă starea conştientizată ar fi alta decât starea proprie, ar urma că subiectul se află simultan în două stări diferite. Pe de altă parte, a conştientiza că mă aflu într-o anumită stare înseamnă a crede că mă aflu în acea stare. Prin urmare, credinţa într-o stare proprie este corectă dacă şi numai dacă mă aflu în acea stare proprie, adică o asemenea credinţă este întemeiată, aşa încât raţionamentele amintite sunt corecte. Dacă cineva ar nega coincidenţa între conţinutul credinţei şi cel al conştientizării, atunci când este vorba de stări proprii, ar trebui să explice cum s-ar ajunge la o asemenea diferenţă şi, în plus, aceasta nu ar afecta relativismul, deoarece în premisă, verbul 'a crede' se poate înlocui, cu acelaşi efect, prin 'a conştientiza': '*X* conştientizează *Fx* $\vdash$ *Fx*' este, astfel, un raţionament corect.

Credinţele însele sunt stări proprii. Prin urmare, '*X* crede că *X* crede *p* $\vdash$ *X* crede *p*' este un raţionament corect, evitându-se regresia la infinit a operatorului credinţei.

Mai mult, nici un subiect nu are acces la stările proprii ale altcuiva. O consecinţă a *teoriei relativităţii* este că fiecare subiect are prezentul propriu: prezentul pentru *X* diferă de prezentul pentru *Y*. În această situaţie, *Y* nu are acces la stările proprii ale lui *X* deoarece nu le poate conştientiza, căci *Y* nu poate

conştientiza decât stările din propriul prezent. Nimeni nu poate şti ce cred eu sau care sunt stările mele proprii, aşa cum nici eu nu pot şti ce crede celălalt sau care sunt stările lui. De aceea, fiecare individ are un domeniu propriu al competenţei; în aceasta constă *principiul toleranţei.*

A accepta teoria relativităţii a lui Einstein înseamnă a accepta relativismul. De aceea, pare ciudat de ce Putnam este de acord cu teoria relativităţii, dar respinge relativismul. Probabil, explicaţia constă în impresia lui Putnam că teoria relativităţii se opreşte la nivelul fizicii, pe când relativismul afectează inclusiv viaţa socială. Dacă ar accepta relativismul, ar însemna să renunţe la teza atât de dragă lui că există standarde obiective pentru binele obştesc şi că democraţia occidentală este cea mai apropiată de asemenea standarde. Dar, o asemenea teză ideologică nu poate fi susţinută cu argumente filosofice. Binele este un termen relativ, de aceea devine vag atunci când încercăm să-l transformăm în termen absolut. De pildă, asupra unei propoziţii precum '*x* este înalt' este dificil să se decidă, dar asta numai pentru că expresia respectivă nu este o propoziţie, adică nu este bine formată. Termenul 'înalt' este relativ, de aceea poate fi folosit numai în contexte precum '*x* este mai (la fel de, mai puţin) înalt decât *y*'; în aceeaşi situaţie sunt termeni precum 'bătrân', 'chel', 'grămadă' etc., aşa explicându-se unele paradoxuri. La fel, propoziţiile de forma '*x* este bun' nu sunt bine formate, (generând ambiguităţi), ci numai '*x* este mai bun decât *y* pentru *z*'. Astfel, nu se poate spune nicicum că democraţia occidentală este un regim politic bun sau mai bun decât altele, ci numai că este un regim politic mai bun decât alte regimuri pentru *unii*, pe când pentru *alţii* poate fi mai rău. Teza că lumea poate fi schimbată într-o lume mai bună este eronată. Orice schimbare este bună pentru unii şi rea pentru alţii. Chiar crima este bună pentru unii, altfel cine ar mai comite crime, cine ar dori, cu tot dinadinsul, să-şi provoace suferinţă sau să-şi facă rău sieşi?

De pildă, în ciuda eforturilor sale, Putnam nu reuşeşte să ne convingă că regalitatea de drept divin ar fi mai puţin raţională decât regimul democratic american. Nici una dintre ele nu rezultă în urma unor raţionamente corecte. Ba chiar, din această

perspectivă, dacă ar exista Dumnezeu şi dacă biserica avea dreptate în acea vreme, regalitatea de drept divin ar fi raţională. Nu există, însă, nici un temei, în afara unei decizii arbitrare, pentru ca instituţii alese prin vot majoritar să deţină dreptul de a decide inclusiv asupra celor care nu le-au votat sau au votat împotrivă. În alte locuri, Putnam însuşi susţine că majoritatea nu are nici un temei să se impună în faţa minorităţii, uneori un singur om deţine adevărul mai bine decât toţi ceilalţi. Am putea spune că regalitatea de drept divin se sprijină pe un raţionament deductiv (problematic este adevărul premiselor) iar regimul democratic reprezentativ se sprijină pe un raţionament inductiv (de la o parte, nu întotdeauna majoritară, la toţi), care, din capul locului, este eronat logic.

Un alt exemplu prin care Putnam încearcă să ne convingă că 'binele' este un termen absolut, nu relativ, este legat de preferinţa pentru îngheţata de vanilie sau pentru cea de ciocolată. Aici, Putnam, în ura sa faţă de relativism, ajunge să abandoneze analiza logică în favoarea psihologismului. Teza lui Putnam este că, pentru a explica de ce lui Smith nu-i place vanilia pe când lui Jones îi place, trebuie să presupunem că acelaşi eşantion de vanilie are simultan două gusturi diferite, unul simţit de Smith (V_S), care este neplăcut (sau rău în sine) şi altul simţit de Jones (V_J) care este plăcut (sau bun în sine). O asemenea teză contravine principiilor logicii care interzic ca un lucru să se afle în două stări opuse în acelaşi timp. Pe de altă parte, dacă gusturile sunt, ele însele, plăcute sau neplăcute, ar însemna că plăcutul sau neplăcutul poate fi simţit cu limba. În acest caz, fiziologii ar fi trebuit să identifice care sunt papilele gustative responsabile pentru senzaţia de plăcere şi pe ce parte a limbii se află, aşa încât să ştim cum să abordăm bucatele pentru a simţi o plăcere continuă. De fapt, trebuie să se accepte că vanilia are un singur gust, atât pentru Smith, cât şi pentru John care nu are a face cu plăcerea, ci numai cu anume tipuri de reacţii fizico-chimice. Sentimentul de plăcere este dependent de starea în care se află degustătorul. Chiar lui Smith, în unele împrejurări, ar putea să-i placă vanilia, iar în altele să o respingă, aşa încât nu se poate

susţine că există un gust al vaniliei legat de Smith şi altul legat de Jones. Vedem că, pentru a explica fenomenele privind gustul, trebuie să ţinem seama de *relaţia* dintre degustător şi degustat, adică de perspectiva relativistă.

Totuşi, s-ar putea construi predicate monare plecând de la predicate n-are prin valorizarea a n–1 variabile. De pildă, predicatul monar 'x este înalt' poate fi definit prin 'x este înalt' $=_{df}$ 'x este mai înalt decât a'; bunăoară, am putea conveni să numim un om înalt dacă are mai mult de 1,80 m în înălţime (sau altă convenţie). Pentru 'x este bun', s-ar putea adopta definiţia: 'x este bun' $=_{df}$ 'x este mai bun decât b pentru s'. Dar, spre deosebire de prima definiţie, a doua este inaplicabilă. În cazul primei definiţii, se ajunge la o relaţie *externă* – oricine poate măsura înălţimea unui obiect dacă se află într-o relaţie potrivită faţă de acesta, – în schimb, nimeni nu poate măsura cu cât este mai bun un lucru decât altul pentru cineva, pentru că aici intervin stări interne, care nu pot fi constatate cu ajutorul simţurilor. O definiţie a binelui precum cea de mai sus, îl transformă pe s într-un arbitru absolut pentru ce este bine şi ce este rău; noi ceilalţi, ar trebui să-l consultăm, să-l ascultăm şi să ne supunem lui necondiţionat.

Dacă trecerea de la propoziţiile de credinţă la propoziţiile de stare este posibilă pentru stările proprii, nu acelaşi lucru se întâmplă pentru stări ale altor obiecte. Nu există acces direct la stările altui obiect; datorită faptului că obiectele sunt în afara mea, înseamnă că stările acestora au avut loc în trecut, ori trecutul nu mai există (teoria relativităţii, pe care Putnam o acceptă, interzice interacţiunile instantanee între lucruri separate spaţial). Ar însemna să posed capacităţi miraculoase dacă aş avea acces la stări care nu există. De aceea, stările mele care sunt determinate de interacţiunea cu alte obiecte sunt suficiente pentru a induce stări proprii, dar nu sunt necesare. Are loc: 'X simte că a este roşu $\vdash$ X simte roşu', dar nu are loc: 'X simte roşu $\vdash$ X simte că a este roşu' (s-ar putea ca senzaţia de roşu să fie provocată de b sau să fie o iluzie etc.). Cu alte cuvinte, premisa ultimei inferenţe nu este suficientă pentru a detaşa

concluzia, ci trebuie adăugată încă o premisă, respectiv, că *X* este într-o anumită relaţie cu *a* (*Rxa*). În acest caz, raţionamentul devine: '*X* crede că *X* simte că *a* este roşu; *Rxa* ⊢ *X* simte că *a* este roşu'. Astfel, pentru ca inferenţa să aibă loc, este nevoie ca subiectul să se afle într-o relaţie potrivită cu obiectul, ori aceasta este o relaţie externă, care poate fi stabilită şi de alt observator. Dacă *X* afirmă că el vede că *a* este roşu, *Y* poate calcula dacă *X* minte sau în ce relaţie se află cu *a*. Se cunoaşte povestea celui care, încercând să scape de acuzaţia că a trecut cu autoturismul pe culoarea roşie a semaforului, a invocat efectul Döppler, care l-a pus într-o anumită relaţie cu semaforul, aşa încât a văzut semaforul verde, dar, prin calcul, se poate constata că, în acest caz, a depăşit viteza legală.

Tocmai acest lucru îl afirmă teoria relativităţii a lui Einstein: doi observatori care se află în relaţii diferite faţă de acelaşi obiect, sunt îndreptăţiţi să creadă propoziţii de observaţie (măsurare) diferite cu privire la el. Dacă un observator afirmă sincer 'Observ (măsor) că *a* are 1 m lungime', iar un altul 'Observ (măsor) că *a* are 1,2 m lungime' nu înseamnă, neapărat, că unul dintre ei se înşeală. Cele două propoziţii nu sunt în opoziţie deoarece au referinţă diferită (referinţa fiind observatorii, nu obiectul). Dacă ambii au dreptate sau nu, se poate verifica de către unul dintre ei sau de către un terţ, cu ajutorul *transformărilor lui Lorentz* care permit unui observator, când cunoaşte relaţia celuilalt observator cu obiectul, să stabilească credinţa îndreptăţită a celuilalt. Prin urmare, trebuie subliniat încă odată, nu se poate, în mod consistent, să se susţină teoria relativităţii a lui Einstein, dar să se respingă relativismul, aşa cum procedează Putnam.

Totuşi, de la propoziţii de observaţie, cum sunt cele de mai sus, nu se poate trece la propoziţii obiective (care se referă la obiect). Raţionamentul '*X* crede că *a* are 1m lungime; *Rxa* ⊢ *a* are 1 m lungime' nu este corect. Dacă s-ar accepta asemenea raţionamente, s-ar ajunge la contradicţie deoarece premisele următoarelor raţionamente pot fi adevărate împreună:

X observă că *a* *Y* observă că

are 1 m lungime.	*a* are 1,2 m lungime.
Rxa	*Rıya*
***a* are 1 m lungime'	***a* are 1,2 m lungime'

Nu se poate să fie adevărate atât '*a* are 1 m lungime' cât şi '*a* are 1,2 m lungime'. Pentru a ieşi din impas, se adoptă *convenţii sociale*: starea lui *a* este stabilită de un observator aflat într-un anumit raport cu obiectul. Numai observaţiile date de o anumită relaţie *R* sunt luate în considerare pentru a stabili starea obiectului. De pildă, deşi tăblia mesei (pentru a relua un exemplu a lui Putnam) este observată ovală sub un anumit unghi, nu se poate spune că 'Tăblia mesei este ovală' dacă unghiul de observaţie nu este de 90°. Dacă sub acest unghi, stabilit prin convenţie, masa este observată rotundă, atunci cine spune 'Tăblia mesei este ovală' se află în eroare, în ciuda observaţiei. Dacă masa este observată sub un unghi oarecare, cunoscând unghiul convenţional, observatorul poate calcula care este forma *reală* a mesei. Prin urmare, *realitatea* este urmarea unor convenţii sociale. S-ar putea adopta o altă convenţie, de pildă, forma mesei să fie cea observată sub un unghi de 45°, când afirmaţiile acceptate în comunitate despre masă s-ar modifica. Propoziţia 'Apa fierbe la 100°C' este adevărată, chiar dacă cineva, aflat la mare altitudine observă că apa nu fierbe la 100°C, dar, o populaţie care ar trăi pe crestele munţilor, ar putea foarte bine să adopte o altă convenţie, în cadrul căreia 'Apa fierbe la 100°C' să fie falsă. Prin urmare, inclusiv propoziţiile obiective, care descriu *realitatea* sunt relative la convenţiile unei anumite comunităţi. Rolul oamenilor de ştiinţă este să cunoască aceste convenţii, să le aplice şi să le transmită. Domeniul propoziţiilor obiective este singurul în care comunitatea este îndreptăţită să impună anumite propoziţii ca adevărate, pe când, pentru propoziţiile subiective, nu există autoritate externă justificată. Orice autoritate care încearcă să impună cuiva propoziţii subiective, de pildă, despre ce este bine sau rău pentru cineva, este abuzivă. Chiar şi relativismul este relativ, căci, în funcţie de propoziţie, autoritatea îndreptăţită să o susţină este mereu alta.

Ajungem, astfel, la a patra obiecţie adusă relativismului. Cele arătate mai sus, conduc la *principiul toleranţei*: 'Oricine poate ajunge la adevăr'. Chiar în cazul propoziţiilor obiective, dacă cineva cunoaşte convenţia socială asupra propoziţiei, respectiv, dacă acela cunoaşte relaţia standard faţă de obiect, se poate plasa el însuşi în acea relaţie cu obiectul şi poate stabili adevărul sau falsitatea propoziţiei obiective. A respinge relativismul, înseamnă a respinge principiul toleranţei, ajungând la (sub condiţia de a respinge şi scepticismul): 'Numai unii pot cunoaşte adevărul'. Opusul relativismului este teza că există anumite persoane care deţin adevărul, de pildă prin revelaţie sau în alt mod, iar ceilalţi nu pot să-l cunoască direct, aşa încât, trebuie să asculte acele persoane şi să se supună lor necondiţionat. Păcat că Putnam, adversar al relativismului, nu ne dezvăluie care ar fi acele persoane privilegiate.

În acest caz, numai privilegiaţii sunt îndreptăţiţi să vorbească cu sens, ceilalţi produc numai zgomot sau sunt *redundanţi*, adică repetă ceea ce rostesc privilegiaţii (aşa cum se întâmplă de multe ori în filosofia din ţara noastră). Respingerea relativismului transformă pe cei mai mulţi oameni în *supuşi*, ei nu pot fi *cetăţeni*. În schimb, dacă acceptăm relativismul şi principiul toleranţei, rezultă că oricine poate avea dreptate, oricine poate vorbi cu sens, oricine *merită* să fie ascultat, prin urmare, nu este adevărat că relativismul conduce la desconsiderarea semenilor şi la transformarea lor în fiinţe inferioare, ci dimpotrivă, este singura atitudine care le recunoaşte demnitatea.

Ionel Narița

Creiere în container

O furnică se târăşte pe un petic de nisip. În urma sa rămâne o urmă. Din pură întâmplare, furnica desenează o figură care arată ca o caricatură a lui Winston Churchill. Oare a reuşit furnica să redea imaginea lui Winston Churchill, să realizeze o imagine care să-l *redea* pe Churchill?

Cei mai mulţi vor spune, după o scurtă gândire, că nu este aşa. Furnica, de fapt, nu l-a văzut niciodată pe Churchill, sau măcar vreo imagine de-a lui Churchill şi nu a avut intenţia să-l înfăţişeze pe Churchill. Pur şi simplu, ea a trasat o linie (în mod neintenţionat), pe care *noi* o vedem ca o reprezentare a lui Churchill.

Putem exprima acest fapt spunând că linia nu este

'ea însăşi' o reprezentare[1] a ceva, mai degrabă decât a orice altceva. În mod similar (chiar dacă de o manieră mai complicată) prezenţa trăsăturilor lui Winston Churchill nu este suficientă pentru a face ca o linie să-l reprezinte sau să se refere la Churchill. Nici nu este necesar: în comunitatea noastră literele tipărite 'Winston Churchill' sau cuvintele rostite 'Winston Churchill', cât şi multe alte lucruri sunt utilizate pentru a-l reprezenta pe Churchill (deşi nu în mod pictural), nu au nici o asemănare cu Churchill aşa cu au imaginile – chiar dacă sunt alcătuite dintr-o simplă linie. Dacă *similaritatea* nu este necesară sau suficientă pentru a face ceva să reprezinte altceva, atunci *ce* ar putea fi necesar şi suficient pentru un asemenea scop? Cum anume ar putea un lucru să reprezinte (sau să 'stea pentru' etc.) un lucru diferit?

La prima vedere, răspunsul pare uşor. Să presupunem că furnica l-a văzut pe Winston Churchill şi să presupunem că are inteligenţa şi abilitatea de a-l desena. Să presupunem că ea produce această caricatură în mod *intenţionat*. În acest caz, linia îl va reprezenta pe Churchill.

Pe de altă parte, să presupunem că linia are forma WINSTON CHURCHILL. Mai presupunem că aceasta se întâmplă pur accidental (fără să ţinem seama de caracterul foarte improbabil). În acest caz, forma scrisă

[1] În această carte termenii 'reprezentare' şi 'referinţă' se referă întotdeauna la o relaţie între cuvinte (sau alte tipuri de semne, simboluri sau reprezentări) şi ceva care există în fapt (nu este un simplu 'obiect al gândirii'). Există un sens pentru 'referă' în care eu pot să mă 'refer' la ceea ce nu există; nu acesta este sensul în care utilizăm acest termen aici. Un cuvânt mai vechi pentru ceea ce numesc 'reprezentare' sau 'referinţă' este *denotare*. În al doilea rând, urmez obiceiul logicienilor moderni de a folosi 'există' cu înţelesul 'există în trecut, prezent sau viitor'. Astfel, Wiston Churchill 'există' şi putem să ne 'referim la' sau să 'îl reprezentăm pe' Winston Churchill, deşi nu mai trăieşte.

WINSTON CHURCHILL *nu* îl reprezintă pe Churchill, deşi aceeaşi formă scrisă îl reprezintă aproape întotdeauna atunci când apare într-o carte din zilele noastre.

Prin urmare, am determinat un factor necesar pentru a reprezenta: *intenţia.*

Dar, pentru a putea avea intenţia de a-l *reprezenta* pe Churchill, chiar prin limbajul particular (cuvintele 'Winston Churchill' rostite în mintea mea şi nu cu voce tare), ar trebui să fim în măsură *să gândim* la Churchill mai întâi. Dacă liniile în nisip, zgomotele etc. nu pot reprezenta ceva 'prin ele însele', atunci cum se întâmplă că pot fi gândite ca reprezentând ceva? Sau ele pot? Cum am putea dobândi şi surprinde ceva care ne este exterior?

În trecut, unii filosofi au folosit acest tip de consideraţii drept dovadă că mentalul, *în mod esenţial*, este de natură *nonfizică.* Argumentul este simplu: ceea ce spunem despre curba trasată de furnică se aplică oricărui obiect fizic. Nici un obiect fizic nu ar putea să se refere la un obiect fizic mai degrabă decât la altul prin el însuşi; neîndoielnic, *gândurile din minte* reuşesc să se refere la ceva mai curând decât la altceva. Prin urmare, gândurile (şi mentalul) sunt în mod esenţial diferite prin natura lor de obiectele fizice. Gândurile au caracteristica *intenţionalităţii* – ele se pot referi la cu totul altceva; nimic fizic nu are 'intenţionalitate', care este derivată dintr-o anume folosire a obiectelor fizice de către minte. Sau aşa se afirmă. Se merge prea departe; postularea unor puteri misterioase ale minţii nu rezolvă nimic. Dar problema rămâne: cum sunt posibile intenţionalitatea şi referirea?

Teoriile magice ale referirii

Am văzut că 'schiţa' furnicii nu are nici o conexiune necesară cu Winston Churchill. Simplul fapt că aceasta 'se aseamănă' cu Churchill nu o transformă într-o imagine propriu-zisă, sau într-o reprezentare a lui Churchill. Numai dacă furnica este una inteligentă (ceea ce nu este cazul) şi îl cunoaşte suficient pe Churchill (ceea ce nu se întâmplă) curba trasată este o imagine sau o reprezentare a acestuia. Unii primitivi cred că unele reprezentări (în speţă *numele*) au o conexiune necesară cu purtătorii lor; că deţinerea 'adevăratului nume' al cuiva sau a ceva oferă putere asupra aceluia. Sursa acestei puteri este *conexiunea magică* între nume şi purtătorul numelui; dacă se acceptă că numele au *numai* o legătură contextuală, contingentă şi convenţională cu purtătorii săi, este dificil de înţeles cum anume cunoaşterea numelui ar avea vreo însemnătate magică.

Ce este valabil pentru imaginile fizice este valabil şi pentru imaginile mentale, ca şi pentru reprezentările mentale în general; reprezentările mentale nu sunt legate în mod mai necesar de ceea ce reprezintă decât cele fizice. Presupunerea contrară este o urmă a gândirii magice.

Se poate că *imaginile* mentale sunt cele mai la îndemână. (Se pare că primul filosof care a surprins marea importanţă a acestora, chiar dacă nu a fost primul care le-a studiat, a fost Wittgenstein). Să presupunem că undeva există o planetă pe care au evoluat oameni (sau au fost aduşi de către extratereştri, sau cum doriţi). Să presupunem că aceşti oameni sunt întru totul asemenea nouă, numai că nu au văzut *arbori*. Să mai presupunem că niciodată nu şi-au imaginat arbori (poate că viaţa vegetală există pe

planeta lor numai sub formă de mucegai). Să presupunem că într-o zi, în mod întâmplător, un desen cu un arbore este aruncat dintr-o navă spaţială care se află în trecere, fără a avea vreun contact cu locuitorii. Să ne imaginăm uimirea lor la vederea desenului. Ce o fi asta? Apar toate variantele de speculaţii: o clădire, un baldachin, sau un animal de un anume fel. Dar să admitem că nu vor ghici adevărul.

Pentru *noi* desenul este reprezentarea unui arbore. Pentru semenii noştri de pe planeta îndepărtată, desenul reprezintă doar un obiect straniu, de o natură şi funcţie necunoscută. Să presupunem că ei au o imagine mentală aidoma imaginii mele mentale provocate de desenul respectiv. Imaginea lor mentală nu este *reprezentarea unui arbore*. Ea este numai reprezentarea unui obiect straniu (oricare ar fi) pe care desenul misterios îl reprezintă.

Unii ar putea încă argumenta că imaginea mentală este *în fapt* o reprezentare a unui arbore numai dacă desenul care cauzează imaginea mentală a fost iniţial reprezentarea unui arbore. Există un lanţ cauzal de la arbori la imaginea mentală, chiar dacă este unul foarte complicat.

Dar chiar lanţul cauzal poate fi imaginat ca absent. Să presupunem că acel 'desen al unui arbore' pe care l-a aruncat nava spaţială nu a fost chiar un desen al unui arbore, ci rezultatul accidental al unor pete de culoare. Chiar dacă ar arăta exact ca desenul unui arbore, el este, de fapt, desen al arborelui în aceeaşi măsură în care 'caricatura' furnicilor este o reprezentare a lui Churchill. Am putea să ne imaginăm că nava spaţială provine de pe o planetă pe care nu sunt cunoscuţi arborii. Chiar într-o asemenea situaţie, oamenii ar avea o imagine mentală echivalentă calitativ cu imaginea mea a unui arbore, dar ei nu ar avea o imagine care să reprezinte un arbore mai mult decât un alt lucru.

Acelaşi fenomen se petrece şi cu *cuvintele*. Un text scris ar putea fi o descriere perfectă a unui arbore, dar dacă ar fi produs întâmplător de o maimuţă care loveşte tastele unei maşini de scris timp de milioane de ani, atunci cuvintele nu s-ar referi la nimic. Dacă cineva ar memora aceste cuvinte şi le-ar repeta în minte fără să le înţeleagă, atunci nu s-ar referi la nimic.

Să ne imaginăm că persoana care repetă mental cuvintele este hipnotizată. Să presupunem că aceste cuvinte sunt în japoneză şi persoana înţelege japoneza. Să presupunem că are senzaţia că înţelege cuvintele. (Aidoma celui care întrebat ce înseamnă cuvintele sale constată că nu reuşeşte să le explice). S-ar putea ca iluzia să fie atât de perfectă de parcă persoana ar fi în legătură cu un telepat japonez! Cu toate acestea, dacă ea nu poate utiliza cuvintele corect în contexte potrivite, nu poate răspunde la întrebări cu privire la ce gândeşte etc., atunci ea nu le înţelege.

Prin punerea cap la cap a acestor exemple ştiinţifico-fantastice putem construi un caz în care cineva gândeşte cuvinte care sunt o descriere a unui arbore într-o anumită limbă *şi* simultan are imaginea lor mentală corectă, dar *nici* nu înţelege cuvintele, *nici* nu ştie ce este un arbore. Am putea accepta în plus că imaginea mentală este cauzată prin împrăştierea vopselei (chiar dacă persoana cu pricina a fost hipnotizată să creadă că are de-a face cu imaginea a ceva familiar – atât doar că, dacă este întrebată, nu poate spune despre ce este vorba). Am putea imagina mai departe că limbajul în care persoana gândeşte nu este nici al hipnotizatorului, nici al hipnotizatului – poate este o coincidenţă că aceste 'enunţuri fără sens', cum le presupune hipnotizatorul, să fie o descriere a arborelui în japoneză. Pe scurt, ceea ce trece prin mintea persoanei, chiar dacă este calitativ identic cu ceea ce se petrece în mintea unui

vorbitor de limba japoneză atunci când *în mod real* se gândeşte la arbori – nu s-ar referi la arbori.

Întrunirea condiţiilor amintite nu este, de bună seamă, posibilă în sensul în care este imposibil ca maimuţele să scrie la tastatură *Hamlet* aşa încât probabilitatea împotriva lor este atât de mare încât nu se vor întâmpla niciodată (cred). Dar nu este imposibil logic, nici chiar imposibil fizic. *S-ar putea* întâmpla (este compatibil cu legile fizicii şi, poate, compatibil cu condiţiile actuale din univers, dacă ar exista suficient de multe fiinţe inteligente pe alte planete). Iar, dacă s-ar întâmpla, ar fi o demonstraţie pentru un adevăr conceptual important; chiar un sistem larg şi complex de reprezentări, atât verbale cât şi vizuale, nu este dat *intrinsec*, construit în interior, conectat magic cu ceea ce reprezintă – o legătură independentă de modul în care a fost cauzată şi de dispoziţiile celui care gândeşte sau vorbeşte. Aceasta este adevărat indiferent dacă sistemul reprezentărilor (cuvinte şi imagini, în cazul de faţă) este realizat în mod fizic – cuvintele sunt scrise sau vorbite, iar tablourile sunt realităţi fizice – sau este realizat numai în minte. Cuvintele gândite şi imaginile mentale nu reprezintă *intrinsec* ceea ce reprezintă.

Cazul creierelor în container

De data aceasta vom avea de-a face cu o altă situaţie ştiinţifico-fantastică discutată mult de filosofi: să ne imaginăm că o persoană (vă puteţi imagina chiar pe dumneavoastră) a fost supusă unei operaţii de către un om de ştiinţă diabolic. Creierul acelei persoane (creierul dumneavoastră) a fost extras din corp şi pus într-un container unde este menţinut în viaţă. Terminaţiile

nervoase sunt conectate la un supercomputer care provoacă creierului iluzia că totul este perfect normal. Toate sunt la locul lor: oamenii, obiectele, cerul etc.; dar în realitate tot ce se întâmplă (poate chiar dumneavoastră) este rezultatul impulsurilor electronice care provin de la computer. Computerul este atât de inteligent încât, dacă persoana în cauză încearcă să ridice mâna, prin feed-back o va face să 'vadă' sau să 'simtă' că are mâna ridicată. Mai mult, modificând programul, omul de ştiinţă diabolic poate cauza victimei sale 'experienţa' (sau halucinaţia) oricărei situaţii sau a oricărui eveniment ar dori acel om de ştiinţă. El ar putea să manipuleze memoria printr-o operaţie pe creier, astfel încât subiectul experimentului să creadă că a fost întotdeauna în acel mediu. Ba chiar, ar putea citi textul de aici şi s-ar amuza de absurdul presupunerii că ar exista un asemenea om de ştiinţă care ar extrage creierele oamenilor din corpuri şi le-ar pune într-un container unde ar fi menţinute în viaţă. Terminaţiile nervoase sunt conectate la un supercomputer care provoacă creierului iluzia că...

Dacă exemplul de mai sus ar face parte dintr-o prelegere de teoria cunoaşterii, scopul urmărit ar fi, bineînţeles, să se ridice clasica problemă a scepticismului cu privire la lumea externă într-o manieră modernă. (*Cum ştim că nu suntem într-o asemenea situaţie?*). Acest exemplu este util şi în ce priveşte problema raportului minte/lume.

În locul unui singur creier în container, am putea imagina că toţi oamenii (chiar toate fiinţele sensibile) au creierele în asemenea incinte (sau sistemele lor nervoase, dacă este vorba de fiinţe cu un sistem nervos minimal). De bună seamă, omul de ştiinţă diabolic ar putea fi în afara experimentului – sau ar putea fi în aceeaşi situaţie? Ar

putea să nu existe nici un om de ştiinţă diabolic, s-ar putea (chiar dacă este absurd) ca universul să constea din asemenea maşinării conectate la creiere sau sisteme nervoase aflate în containere.

De această dată, să presupunem că automatul este programat să genereze o halucinaţie *colectivă*, mai degrabă decât un anumit număr de halucinaţii separate fără legătură între ele. Astfel, când mie mi se pare că vă vorbesc, vouă vi se pare că îmi auziţi vorbele. De bună seamă, nu avem de-a face cu situaţia în care vorbele mele ajung la urechile voastre, – deoarece nu aveţi urechi (în realitate) şi nici eu nu am o gură şi limbă reale. Mai curând, când îmi formulez vorbele, ceea ce are loc este un transfer de impulsuri de la creierul meu la computer, care mă determină să-mi 'aud' propria voce rostind aceste cuvinte, să-mi simt limba mişcându-se etc., dar şi vouă vă provoacă 'auzirea' cuvintelor mele, 'vederea' faptului că vorbesc etc. În acest caz, comunicăm într-un mod efectiv. Eu nu greşesc cu privire la propria mea existenţă (numai în legătură cu existenţa corpului meu şi a 'lumii externe', din afara creierului). Dintr-o anumită perspectivă, nu are importanţă dacă 'întreaga lume' este rezultatul unei halucinaţii colective; ceea ce contează, în ultimă instanţă, este auzirea cuvintelor mele atunci când vă vorbesc, chiar dacă mecanismul nu este cel presupus în mod obişnuit. (Bineînţeles, dacă suntem doi amanţi care fac dragoste, mai curând decât doi oameni care poartă o discuţie, atunci sugestia că sunt implicate numai două creiere ar putea fi ciudată).

Acum voi formula o problemă care pare prostească şi evidentă (cel puţin pentru unii oameni, inclusiv filosofi versaţi), dar a cărei adâncime filosofică va reieşi curând. Să presupunem adevărate cele de mai sus. Am putea, dacă am

fi creiere aflate în containere în modul arătat, *să spunem* sau *să gândim* că existăm?

Doresc să susţin că răspunsul la această întrebare este 'Nu, nu am putea susţine aceasta'. De fapt, doresc să argumentez că supoziţia că nu suntem altceva decât creiere în containere, deşi nu violează nici o lege fizică, şi este perfect consistentă cu întreaga noastră experienţă, nu ar putea fi adevărată. *Nu ar putea fi adevărată* deoarece este, într-un fel, autocontradictorie.

Argumentul pe care am să-l prezint este unul neobişnuit şi mi-au trebuit mai mulţi ani să mă conving că aşa stau lucrurile. Dar, este un argument corect. Ceea ce îl face atât de straniu este conexiunea cu unele dintre cele mai profunde teme filosofice. (Argumentul mi-a venit în minte prima oară când meditam asupra unei teoreme a logicii moderne, teorema 'Skolem-Löwenheim' şi dintr-o dată am văzut legătura între această teoremă şi unele teze din *Investigaţiile filosofice* ale lui Wittgenstein).

O 'supoziţie care se respinge singură' este una pentru care adevărul său implică propria falsitate. De exemplu, să considerăm teza că *toate judecăţile generale sunt false.* Aceasta este o judecată generală. Astfel, dacă este adevărată, atunci trebuie să fie falsă. Deci, este falsă. Despre unele teze se spune că 'se resping singure' dacă *supoziţia pe care se sprijină acele teze* implică propria falsitate. De pildă, 'Eu nu exist' se respinge singură dacă mă am în vedere pe *mine* (pentru orice *mine*). Cu alte cuvinte, cineva poate fi sigur că există, odată ce se gândeşte la asta (după cum a argumentat şi Descartes).

Voi arăta că presupunerea că suntem creiere în containere are tocmai o asemenea proprietate. Dacă putem presupune această teză ca adevărată sau falsă, atunci urmează să arăt că nu este adevărată . Deci nu este adevărată.

Înainte de a-l dezvolta, să vedem de ce un asemenea argument pare atât de ciudat (cel puţin filosofilor care subscriu la teoria corespondenţei în ce priveşte adevărul). Am căzut de acord că situaţia amintită este compatibilă cu legile fizicii, aşa încât ar trebui să existe o lume în care să fim creiere în containere. Cum ar spune filosofii, există o lume posibilă în care toate fiinţele senzitive sunt creiere în containere. (Conceptul de 'lume posibilă' nu trebuie înţeles ca desemnând un loc unde orice presupoziţie, oricât de absurdă, este adevărată). Oamenii din această lume posibilă au exact aceleaşi experienţe pe care le avem *noi*. Ei gândesc aceleaşi gânduri ca şi noi (cel puţin, prin minţile lor trec aceleaşi cuvinte, imagini, forme de gândire etc.). Cu toate acestea, am afirmat că se poate argumenta că nu suntem creiere în containere. Care este acest argument? Şi de ce oamenii din lumea posibilă care sunt în *realitate* creiere în containere nu îl pot formula?

În linii mari, răspunsul este acesta: deşi oamenii din acea lume posibilă pot gândi şi 'spune' orice cuvinte pe care le putem gândi şi spune noi, ei nu pot (afirm eu) *să se refere* la ce ne referim noi. În particular, ei nu pot gândi sau spune că ei sunt creiere în containere (*chiar atunci când gândesc 'noi suntem creiere în containere'*).

Testul Turing

Să presupunem că s-a inventat un computer care poate duce o conversaţie inteligentă cu un interlocutor (asupra a tot atâtea subiecte câte ar putea o persoană inteligentă). Cum am putea decide dacă acel computer are 'conştiinţă'?

Logicianul britanic Alan Turing propune următorul test[2]: fie cineva care duce o discuţie cu un computer şi o conversaţie cu o persoană pe care nu o cunoaşte. Dacă el nu poate spune care este persoana şi care este computerul, atunci computerul are conştiinţă (presupunem că testul se repetă de un număr suficient de ori cu diferiţi interlocutori). Altfel spus, un automat are conştiinţă dacă trece 'testul Turing'. (De bună seamă, conversaţia nu are loc faţă în faţă, odată ce interlocutorul nu este cunoscut ca aparenţă vizuală. Nu este folosită vocea, odată ce vocea mecanică poate fi diferită de vocea umană. Să ne imaginăm mai degrabă o conversaţie prin intermediul textelor la maşina de scris electrică. Interlocutorul scrie la tastatură enunţurile, întrebările etc., iar cei doi parteneri – maşina şi persoana – răspund printr-o tastatură electrică. Pe de altă parte, maşina poate *mintă* – la întrebarea 'Eşti o maşină?', ar putea răspunde prin 'Nu, sunt asistentul de laborator'.

Ideea că acest test stabileşte într-adevăr prezenţa conştiinţei a fost criticată de un număr de autori (care nu sunt ostili în principiu ideii că o maşină ar putea avea conştiinţă). Dar nu asta ne preocupă acum. Aş dori să folosesc ideea generală a testului Turing, ideea unui *test al competenţei prin intermediul dialogului*, pentru un alt scop, şi anume, explorarea noţiunii de *referinţă*.

Să ne imaginăm o situaţie în care problema nu este de a determina dacă partenerul este o maşină sau o persoană, ci mai degrabă, dacă partenerul utilizează cuvintele cu aceeaşi referinţă ca şi noi. Ca mai sus, testul constă în a purta o conversaţie şi, dacă nu apar alte probleme, dacă partenerul 'trece' în sensul că nu poate fi

[2] A.M. Turing, 'Computing Machinery and Intelligence', *Mind* (1950), retipărit în A.R. Anderson (ed.), *Minds and Machines.*

deosebit de cineva despre care ştim dinainte că utilizează acelaşi limbaj cu noi, aceleaşi reguli de referire etc., putem concluziona că subiectul testului se referă la aceleaşi obiecte la care ne referim noi. Prin *Testul Turing pentru referinţă* am în vedere un test Turing cu scopul de a determina existenţa unei singure referinţe (împărtăşită) pentru doi sau mai mulţi interlocutori. Aşa cum despre testul Turing se ridică problema dacă este suficient pentru a stabili conştiinţa, să vedem dacă testul Turing pentru referinţă este suficient pentru referinţa împărtăşită.

Răspunsul este 'nu'. Testul Turing pentru referinţă nu este decisiv. El are un rol practic; dar nu este logic imposibil (deşi foarte improbabil) să existe cineva care să treacă testul Turing pentru referinţă şi să nu se refere la ceva. Urmează, aşa cum vom vedea, că putem extinde observaţia anterioară că un cuvânt (chiar textele în întregime şi discursurile) nu are legătură necesară cu referinţa sa. Chiar dacă am considera nu atât cuvintele propriu-zise, cât regulile care decid ce cuvinte este potrivit să apară în anumite contexte – chiar dacă am avea în vedere, în jargonul informaticienilor, *programele pentru folosirea cuvintelor* – acele programe care se *referă la ceva extralingvistic* încă nu determină referinţa pe care o au cuvintele. Acesta va fi un pas crucial pentru a ajunge la concluzia că creierele-în-containere nu se pot referi la nimic exterior (aşa încât nu pot spune că ele sunt creiere-în-containere.

Să presupunem, de pildă, că mă aflu într-o situaţie Turing (jucând, în terminologia lui Turing, 'jocul imitaţiei'), iar partenerul meu este o maşină. Să presupunem că această maşină este în măsură să câştige jocul. ('Să treacă' testul.) Să ne imaginăm că maşina a fost programată să dea răspunsuri potrivite în limba engleză la enunţuri, întrebări, remarci etc., dar că nu are organe de

simţ (altele decât legătura la tastatura mea electrică) şi nici organe efectoare (altele decât cele care acţionează tastatura electrică). (După cum s-a văzut, Turing nu presupune că posesia organelor de simţ sau a celor efectoare este necesară pentru conştiinţă sau inteligenţă). Să mai presupunem că maşina, nu numai că nu are ochi sau urechi electronice, dar nici nu are instrucţiuni în cadrul programului necesare pentru a juca jocul imitaţiei pentru intrări aidoma organelor de simţ sau pentru a controla un corp. Ce ar trebui să spunem despre o asemenea maşină?

După cum cred, nu am putea să atribuim vreo funcţie de referire unui asemenea dispozitiv. Este adevărat că maşina poate discuta cu multă acurateţe despre, să spunem, peisajul din New England. Cu toate acestea, nu ar putea să recunoască un pom fructifer sau un fruct, un munte sau o vacă, un câmp sau o turlă, dacă ar fi faţă în faţă cu vreuna dintre ele.

Avem de-a face cu un dispozitiv care produce enunţuri ca răspunsuri la enunţuri. Dar nici unul dintre ele nu este legat de lumea reală. *Dacă am pune două asemenea maşini ca parteneri în jocul imitaţiei atunci ele s-ar 'păcăli' una pe alta indiferent dacă restul lumii ar dispare*! Nu există nici o raţiune de a considera discursul unei maşini despre meri ca referindu-se la merii din lumea reală altfel decât 'schiţa' furnicilor s-ar referi la Winston Churchill.

Iluzia referirii, a semnificaţiei, a inteligenţei, este produsă aici de faptul că există o convenţie de reprezentare care ne afectează pe *noi* aşa încât discursul maşinii apare ca referindu-se la meri, peisaje sau la New England. Analog, apare *iluzia* că furnicile au reuşit o caricatură a lui Churchill, din acelaşi motiv. Dar, noi suntem cei capabili să percepem şi să manevrăm, să avem de-a face cu mere sau cu câmpuri. Discursul nostru despre mere sau câmpuri este intim legat prin relaţii *neverbale* cu mere şi câmpuri. Există 'reguli de

intrare în limbaj' care provin din experienţele cu mere pentru enunţuri precum 'Eu văd un măr' şi 'reguli de ieşire din limbaj' care ne ajută să luăm decizii exprimate lingvistic ('Am de gând să cumpăr câteva mere') şi posibilitatea de a acţiona altfel decât vorbim. Dacă asemenea reguli, fie de intrare, fie de ieşire, nu există, nu există nici raţiunea de a considera conversaţia cu o maşină (sau între două maşini, dacă examinăm maşinile în timpul jocului imitaţiei) mai mult decât un joc sintactic. Jocul sintactic *se aseamănă* cu discursul inteligent, asta este limpede; dar numai atât (şi nu mai mult de atât) cât liniile trasate de furnici s-ar asemăna cu o caricatură.

În cazul furnicii, s-ar putea argumenta că furnica ar trasa aceeaşi curbă chiar dacă Winston Churchill nu ar fi existat niciodată. În cazul maşinii, nu s-ar putea dezvolta un argument de acest fel; dacă merele, copacii, clopotniţele şi câmpurile nu existau, atunci, probabil, programatorii nu ar fi produs un astfel de program. Deşi maşina nu *percepe* merele, câmpurile sau clopotniţele, creatorul-designer le-a perceput. Există o *oarecare* conexiune cauzală între maşină şi merele din lumea reală prin intermediul experienţei şi cunoştinţelor creatorului-designer. Cu toate acestea, o asemenea conexiune este prea slabă pentru referire. Nu numai că este posibil logic, deşi este cu totul improbabil, aceeaşi maşină ar *putea* exista chiar dacă merele, câmpurile şi clopotniţele nu ar exista; mai important, maşina este cu totul insensibilă la existenţa continuă a merelor, câmpurilor, clopotniţelor etc. Chiar dacă toate acestea ar *înceta* să existe, discursul maşinii s-ar derula în acelaşi mod. Iată de ce, despre o maşină, nu se poate considera că se referă la ceva.

Aspectul relevant pentru discuţia de aici este că nu există nimic în testul Turing care să determine o maşină programată să joace jocul imitaţiei să facă altceva decât să

joace acest joc, iar o maşină care îl joacă nu se referă mai mult decât o face orice înregistrare pe magnetofon.

Creiere în container (iarăşi)

Să comparăm ipoteticele 'creiere în containere' cu maşinile descrise anterior. Există, de bună seamă, diferenţe importante. Creierele în containere nu au organe de simţ, dar ele au înzestrarea necesară pentru organe de simţ, adică au terminaţiile nervoase corespunzătoare, au intrări ale acestor terminaţii nervoase care figurează în 'programul' creierelor în container în acelaşi mod ca în programul creierelor noastre. Creierele în container sunt *creiere*; adică ele *funcţionează* precum creierele, respectiv, potrivit aceloraşi reguli cu creierele din lumea actuală. Din aceste motive ar fi absurd să negăm că ar avea conştiinţă sau inteligenţă. Dar faptul că ele sunt conştiente sau inteligente nu înseamnă că, prin cuvintele lor, se referă la ceea la ce se referă cuvintele în mod obişnuit. Problema pe care o urmărim este dacă, atunci când discursul lor conţine, să spunem, cuvântul 'arbore', se referă într-adevăr la *arbori*? (Prin opoziţie, de exemplu, cu obiectele din imaginea produsă de maşinile automate).

Pentru a ne clarifica ideile, să considerăm că maşinile automate au apărut datorită unei întâmplări (sau coincidenţe) cosmice (sau, poate, au existat dintotdeauna). În această lume ipotetică, maşinile automate nu au un designer-creator. În fapt, cum am afirmat la începutul acestui capitol, am putea să ne imaginăm că orice fiinţă senzitivă (chiar cu abilităţi senzitive minime) trăieşte în interiorul unui container.

O asemenea presupunere nu ajută deoarece nu există conexiune între *cuvântul* 'arbore' aşa cum este folosit de

către aceste creiere şi arborii actuali. Acestea ar folosi cuvântul 'arbore' aşa cum am arătat, gândind numai gândurile lor, având numai imaginile pe care le au, chiar dacă nu ar exista arbori reali. Imaginile sau cuvintele lor sunt calitativ identice cu imaginile şi cuvintele care reprezintă arbori în lumea *noastră*; dar noi am văzut deja (furnica din nou) că similaritatea calitativă cu ceva care reprezintă un obiect (Winston Churchill sau un arbore) nu este suficientă pentru a asigura reprezentarea. Pe scurt, creierele în container nu se gândesc la un arbore real când au gândul 'există un arbore în faţa mea' deoarece nu există nimic în virtutea căruia gândul 'arbore' să reprezinte arbori actuali.

Dacă acest scenariu vi se pare exagerat, reflectaţi la următorul: am văzut că cuvintele nu se referă cu necesitate la arbori chiar dacă sunt aranjate în secvenţe care sunt identice cu un discurs care (atunci când apare în mintea cuiva) ar fi fără îndoială *despre arborii* din lumea actuală. Nici chiar programul, în sensul de reguli, practici, dispoziţii ale creierului pentru un comportament verbal, nu se referă în mod necesar la arbori şi nu este suficient pentru referirea la arbori prin conexiunile pe care le stabileşte între cuvinte, sau între chei şi răspunsuri lingvistice. Dacă aceste creiere s-ar referi la arborii reprezentaţi (arborii reali din afara containerului), atunci aceasta ar avea loc datorită modului în care 'programul' conectează sistemul lingvistic cu intrările şi ieşirile *non-verbale*. Creierele din container au asemenea intrări şi ieşiri (iarăşi acei nervi senzitivi şi terminaţiile nervoase corespunzătoare!), dar am stabilit că 'datele simţurilor' produse prin maşinile automate nu reprezintă arbori (sau ceva extern) chiar când sunt identice cu imaginea unui arbore. Întocmai cum o pată de culoare s-ar putea asemăna cu o pictură care înfăţişează un arbore, fără *a fi* pictura unui arbore, tot aşa 'datele simţurilor' ar putea fi calitativ identice cu o 'imagine a unui arbore' fără a fi o

imagine unui arbore. Cum ar putea faptul că, în cazul creierelor în container, limbajul care este conectat prin program cu intrările senzoriale care nu reprezintă intrinsec sau extrinsec arbori (sau ceva extern), să se refere la arbori sau altceva extern?

Răspunsul este că nu ar putea. Întregul sistem de date ale simţurilor, şi semnale motoare ale terminaţiilor senzitive, şi gânduri mediate verbal sau conceptual conectate prin 'regulile de intrare ale limbajului' cu datele simţurilor (sau în alt mod) ca intrări şi 'regulile lingvistice de ieşire' spre semnalele motoare ca ieşiri, nu au o legătură mai puternică cu *arborii* decât curba furnicilor cu Winston Churchill. Odată ce am văzut că această *similaritate calitativă* (analogă, dacă doriţi, cu identitatea calitativă) între gândurile creierelor în container şi gândurile cuiva din lumea actuală nu presupune coincidenţa de referire, nu este dificil să susţinem că nu există nici o bază pentru a considera că creierele în container s-ar referi la vreun lucru extern.

Premisele argumentului

Voi formula acum argumentul promis care arată că creierele din container nu pot gândi sau spune că ele sunt creiere în container. Mai rămâne să-l explicităm şi să-i examinăm structura.

Aşa cum am spus, când un creier în container (în lumea în care orice fiinţă sensibilă este şi a fost un creier într-un container) gândeşte 'În faţa mea se află un arbore', gândul său nu se referă la arborii actuali. Potrivit unor teorii pe care le voi discuta, s-ar putea referi la arborii din imagine sau la impulsurile electronice care provoacă experienţa arborilor sau la aspecte ale unui program care este responsabil pentru impulsurile respective. Aceste teorii nu

sunt respinse prin afirmaţia anterioară, deoarece există o conexiune cauzală între folosirea cuvântului 'arbore' în româna-container şi prezenţa arborelui în imagine ori prezenţa impulsurilor electrice de un anumit fel sau anumite caracteristici ale programului maşinăriei. Potrivit acestor teorii, creierul este *îndreptăţit* atunci când gândeşte 'În faţa mea se află un arbore'. Dacă ţinem seama la ce se referă cuvintele 'arbore' şi 'în faţa' în româna-container, presupunând că o asemenea teorie este corectă, atunci condiţiile de adevăr pentru 'În faţa mea se află un arbore', când apare în româna-container, constau în aceea că un arbore din imagine este 'în faţa' a ceea ce înseamnă 'mea' în asemenea împrejurări – adică imaginea – sau că, de la maşina automată vin impulsuri electrice care produc asemenea experienţe, sau poate că acţionează acele trăsături ale maşinii despre care se presupune că produc o experienţă de tipul 'în faţa mea se află un arbore'. Asemenea condiţii de adevăr ar fi, de bună seamă, îndeplinite.

Într-un argument de acest gen, 'container' se referă la containerele din imagine, conform românei-container, sau la ceva legat de acestea (impulsuri electrice ori caracteristici ale unui program), dar nu la containere reale, deoarece folosirea cuvântului 'container' în româna-container nu se află în relaţie cauzală cu containerele actuale (nu luăm în seamă conexiunea cauzală care există între cuvintele din româna-container şi containerul în care se află creierul respectiv. Această conexiune nu este una specifică, aşa cum ar fi aceea dintre cuvântul 'container' şi containere). În mod similar, 'fluid nutritiv' se referă la un lichid din imagine în româna-container, sau la ceva legat de aceasta (impulsuri electrice sau caracteristici ale programului). Urmează că, dacă 'lumea posibilă' a lor este cea efectivă, şi chiar noi suntem creiere în container, atunci, 'noi suntem creiere în container' înseamnă că *noi suntem creierele în container din imagine* sau ceva de

acest fel (dacă ar însemna ceva). Dar o parte a ipotezei că suntem creiere în container este că nu suntem creiere în container în imagine (de pildă, nu este o halucinaţie că suntem creiere în container). Prin urmare, dacă suntem creiere în container, atunci propoziţia 'Noi suntem creiere în container' spune ceva fals (dacă spune ceva). Pe scurt, dacă suntem creiere în container, atunci 'Noi suntem creiere în container' este falsă. În concluzie ea este (în mod necesar) falsă.

Presupunerea că o asemenea afirmaţie poate fi adevărată apare din combinarea a două erori: 1) *posibilitatea fizică* este luată prea în serios; şi 2) operarea, fără să ne dăm seama, cu o teorie magică a referinţei, o teorie în care anumite reprezentări mentale se referă cu necesitate la lucruri şi tipuri de lucruri externe.

Există o 'lume posibilă fizic' în care noi suntem creiere în container – ce înseamnă aceasta dacă nu posibilitatea de a exista o *descriere* a unei stări de lucruri care este compatibilă cu legile fizicii? Aşa cum există o tendinţă în cultura noastră (care datează din secolul al XVII-lea) să considerăm *fizica* drept metafizica noastră, adică, să considerăm ştiinţele exacte drept descrierea 'ultimei şi adevăratei înfăţişări a universului', în acelaşi mod, ca o consecinţă imediată, apare tendinţa de a considera 'posibilitatea fizică' drept ceea ce ar putea fi cazul în mod real. Adevărul este adevărul fizic; posibilitatea fizică este posibilitatea; necesitatea este necesitatea fizică etc. Dar, cum am văzut, chiar dacă numai în cazul unui exemplu imaginar, acest punct de vedere este eronat. Existenţa unei 'lumi fizic posibile' în care creierele sunt în containere (şi vor fi întotdeauna) nu înseamnă că aceasta ar putea exista în realitate, adică creierele în containere să fie posibile în mod efectiv. Aşadar, regulile pentru o asemenea posibilitate nu sunt fizice, ci *filosofice*.

Unii filosofi, dornici, în acelaşi timp, să susţină, cât şi să minimizeze anumite afirmaţii (cazul tipic al gândirii anglo-americane din secolul al XX-lea), ar spune: 'De bună seamă, aţi arătat că unele lucruri care aparent sunt posibilităţi fizice sunt imposibilităţi *conceptuale*. Ce este atât de ciudat aici?'

Argumentul meu poate fi descris ca privind 'conceptualul'. A descrie activitatea filosofică drept căutare a adevărurilor 'conceptuale' mi se pare că este acelaşi lucru cu *cercetare privind sensul cuvintelor*. Ori noi nu dorim de loc un asemenea demers.

Pur şi simplu ne propunem să investigăm *precondiţiile* pentru a *gândi despre*, *a reprezenta*, *a se referi la* etc. Am cercetat aceste precondiţii *fără* a investiga sensurile unor cuvinte şi fraze (cum ar face un lingvist, de pildă), ci prin *raţionament a priori*. Nu în vechiul înţeles 'absolut' (chiar dacă am afirmat că teoriile magice ale referinţei sunt *a priori* eronate), ci în sensul de cercetare a ceea ce este *raţional* posibil *presupunând* anumite premise generale, sau făcând ipoteze teoretice foarte largi. Un asemenea procedeu nu este nici 'empiric' nici 'a priori', ci are elemente ale ambelor căi de investigare. În ciuda slăbiciunilor modului meu de a proceda şi a dependenţei de presupunerile care pot fi calificate 'empirice' (cum ar fi presupunerea că mintea nu are acces la lucruri sau proprietăţi externe dacă nu intervin simţurile), metoda mea are strânse afinităţi cu ceea ce I. Kant numeşte investigaţie 'transcendentală'; deoarece este o investigaţie, repet, a *precondiţiilor* referinţei şi în consecinţă ale gândirii – precondiţii construite conform naturii minţilor noastre, deşi nu (după cum spera Kant) cu totul independent de influenţe empirice.

Una dintre premisele argumentului este clară: teoriile magice ale referinţei sunt eronate atât din punct de

vedere mental, cât şi fizic. Cealaltă premisă este că nimeni nu se poate referi la un anumit tip de lucruri (de pildă, *arbori*), dacă nu există nici o interacţiune cauzală cu vreunul dintre ele[3], sau cu lucruri care servesc la descrierea lor. Dar, de ce trebuie să fie acceptate aceste premise? După cadrul larg pe care l-am dezvoltat până acum, a sosit momentul ca ele să fie examinate mai îndeaproape.

Raţiunile pentru a respinge conexiunea necesară dintre reprezentări şi referenţii lor

Am arătat anterior că unii filosofi (cel mai important fiind Brentano) au atribuit minţii o putere numită 'intenţionalitate', menită să explice *referirea*. Eu nu o accept ca soluţie. Dar cine îmi dă acest drept? Sunt eu, poate, prea grăbit?

Aceşti filosofi nu afirmau că putem gândi despre lucrurile externe sau despre proprietăţile lor în absenţa unor reprezentări. Argumentul pe care l-am dat mai înainte comparând datele simţului văzului cu 'desenul' furnicii (argumentarea prin povestirea ştiinţifico-fantastică despre 'imaginea' unui arbore obţinută prin stropire cu o pensulă şi care se prezintă simţurilor asemenea 'imaginii vizuale a unui arbore', fără a fi însoţită de vreun *concept* al

[3] Dacă creierele în container ar avea conexiuni cauzale cu, să spunem, arborii *din viitor*, atunci poate că este posibilă referirea *acum* la arbori prin descrierea 'lucrurile la care vreau să mă refer prin "arbori" sunt cutare sau cutare din viitor'. Dar avem în vedere cazul în care creierele din container nu vor intra *niciodată* în conexiune cauzală cu ceva din afara containerului, deci nici cu arborii viitori etc.

unui arbore) ar fi acceptat ca dovadă că *imaginile* nu servesc întotdeauna referirii. Dacă există reprezentări mentale care se referă în mod necesar (la lucruri externe) ele trebuie să fie de tipul *conceptelor* şi nu de tipul imaginilor. Dar ce sunt *conceptele*?

Dacă apelăm la introspecţie nu percepem 'conceptele' prezente în minţile noastre. Prin analiza fluxului gândirii ajungem la cuvinte, imagini, senzaţii, simţăminte. Când exprim cu voce tare gândurile mele, nu gândesc de două ori. Eu îmi aud propriile cuvinte aşa cum le auziţi şi voi. Există diferenţe între a rosti cuvinte în care cred şi cuvinte în care nu cred (uneori, când sunt nervos, sau în faţa unei audienţe ostile, mă comport de parcă aş minţi, deşi spun adevărul); de asemenea, există diferenţe când rostesc cuvinte pe care le înţeleg şi când rostesc cuvinte pe care nu le înţeleg. Dar îmi pot imagina fără dificultate că cineva crede tocmai aceste cuvinte (în sensul de a spune ce are în minte) şi are simţământul înţelegerii, asertării etc. pe care îl am eu, şi îmi dau seama un minut mai târziu (sau am fost trezit de un hipnotizator) că nu am înţeles ce a trecut prin mintea lui, dacă el a înţeles limbajul în care au fost rostite cuvintele. Nu afirm că este tocmai aşa; pur şi simplu spun că nu este nimic neimaginabil într-o asemenea situaţie. Aceasta nu arată despre concepte că *sunt* cuvinte (sau imagini, senzaţii etc.), ci a atribui cuiva un 'concept' sau un 'gând' este altceva decât a atribui vreo 'prezenţă' mentală, ori vreo entitate accesibilă prin introspecţie. Conceptele nu sunt prezenţe mentale care se referă la obiecte externe pentru motivul decisiv că ele nu sunt prezenţe mentale în nici un caz. Conceptele sunt semne folosite într-un

anume fel; semnele pot fi publice sau private, entităţi mentale sau fizice, dar chiar dacă semnele sunt 'mentale' sau 'private', semnul, în afara utilizării sale ca semn, nu este un concept. Semnele, prin ele însele, nu au funcţia referinţei.

Acest fapt poate fi pus în evidenţă printr-un experiment mental foarte simplu. Să presupunem că sunteţi asemenea mie şi nu puteţi deosebi un ulm de un fag. Presupunem că ne referim la acelaşi lucru prin 'ulm' şi că mulţimea tuturor fagilor este extensiunea lui 'fag' (adică mulţimea lucrurilor despre care 'fag' poate fi predicat cu adevărat). Oare diferenţa între referinţa lui 'ulm' şi referinţa lui 'fag' este dată de diferenţa dintre *conceptele* corespunzătoare acestora? Conceptul meu de 'ulm' este acelaşi cu conceptul meu de 'fag' (sunt nevoit să recunosc). (Aceasta arată că determinarea referinţei este socială şi nu individuală, pe de o parte; nu ne putem compara cu un expert care *poate* distinge ulmii de fagi). Dacă cineva s-ar strădui să susţină că diferenţa între cele două referinţe din vorbirea *mea* este determinată de starea psihologică, să-l rugăm să-şi imagineze un Pământ geamăn în care cei doi arbori sunt schimbaţi între ei. Pământul geamăn seamănă foarte mult cu Pământul; de fapt, cu excepţia faptului că 'ulm' şi 'fag' sunt schimbate între ele, cititorul poate presupune că cele două planete sunt identice. Să presupunem că pe Pământul geamăn am un *tiz* ale cărui molecule sunt identice cu ale mele. Dacă sunteţi un dualist, atunci veţi accepta că tizul meu gândeşte aceleaşi gânduri verbalizate ca şi mine, are aceleaşi date ale simţurilor, aceleaşi dispoziţii etc. Ar fi absurd să se presupună că stările lui psihologice sunt

diferite de ale mele: cu toate acestea, cuvântul său 'ulm' stă pentru *fagi*, iar cuvântul meu 'ulm' stă pentru ulmi. (În acelaşi fel, pe un Pământ geamăn, dacă apa este un lichid diferit, să spunem XYZ şi nu H_2O, atunci 'apă' reprezintă un lichid diferit când este folosit pe Pământul geamăn faţă de utilizarea sa pe Pământ etc.). Contrar doctrinei acceptate încă din secolul al XVII-lea, *sensurile nu se află în cap*.

Am văzut că a poseda un concept nu înseamnă a poseda imagini (să spunem a unor arbori – sau chiar imagini, 'vizuale' sau 'acustice', ale enunţurilor, ori ale unor discursuri, fără deosebire) deoarece este posibil să posedăm un sistem de imagini fără a avea *abilitatea* de a folosi enunţurile în mod corect (considerând atât factorii lingvistici – pe care i-am amintit anterior – cât şi cei nelingvistici ca determinând situaţiile caracteristice). Un om poate avea orice imagine am dori şi încă să nu fie în stare să indice un arbore, chiar dacă ar fi într-o pădure. El ar putea avea chiar imaginile pe care le presupunem că le are şi să nu fie în stare să facă ceea ce i se cere. O imagine, dacă nu este însoţită de abilitatea de a acţiona într-un mod specific, este doar o *reprezentare* şi a acţiona conform unei reprezentări este ea însăşi o abilitate pe care o putem avea sau nu. (Un om s-ar putea reprezenta pe el însuşi indicând spre un arbore, de dragul de a lua în considerare orice posibilitate logică; el ar arăta spre un arbore după ce cineva ar produce o secvenţă de sunete – care pentru el nu ar avea sens – secvenţa de sunete 'te rog să arăţi un arbore'). Nici într-o asemenea situaţie nu ar şti ce înseamnă a arăta un arbore şi nu ar *înţelege* 'arată un arbore'.

Am cercetat abilitatea de a folosi anumite enunţuri drept criteriu pentru posedarea pe deplin a unui concept, dar acesta poate fi nuanţat cu uşurinţă. Am putea accepta simbolisme care nu sunt alcătuite din cuvinte ale limbajului natural sau am putea accepta fenomene mentale cum ar fi imagini sau alte tipuri de evenimente interne. Aici este esenţială prezenţa unei anume complexităţi, abilitatea de a combina elementele unele cu altele etc., aşa cum sunt enunţurile din limbajul natural. Astfel, chiar dacă o reprezentare particulară – cum ar fi o licărire albastră – ar putea fi privită ca prezenţă internă pentru demonstraţia teoremei numerelor prime în cazul unui matematician, încă nu ar fi nimerit să se susţină acest lucru dacă licărirea nu ar fi decompozabilă în paşi separaţi şi conexiuni logice. Dar, nu contează ce tip de fenomene interne ar fi considerate ca *expresii* posibile pentru gânduri, argumente precum cel anterior vor arăta că nu fenomenele însele constituie înţelegerea, ci mai degrabă abilitatea gânditorului de a *folosi* aceste fenomene, de a produce fenomenul potrivit în circumstanţa potrivită.

Argumentul anterior este o variantă abreviată a argumentului lui Wittgenstein din *Investigaţii filosofice*. Dacă este corect, atunci tentativa de a înţelege gândirea prin aşa-numita 'investigaţie fenomenologică' este fundamental eronată, deoarece ceea ce fenomenologii numesc *expresie* internă a gândirii nu este o *producere* ci o *abilitate*. Exemplul precedent cu subiectul care pretinde că gândeşte în japoneză (inducând în eroare un telepat japonez) este suficient pentru a dovedi eroarea abordării fenomenologice în problema *înţelegerii*. Chiar dacă ar

exista calităţi ce pot fi puse în evidenţă prin introspecţie prezente numai atunci când are loc înţelegerea (aceasta ar fi o introspecţie falsă, de fapt), încă aceste calităţi nu ar fi suficiente pentru a susţine că omul respectiv ar înţelege vreun cuvânt în japoneză.

Pe de altă parte, cred că situaţia în care cineva nu are nici un fel de 'monolog interior' este posibilă. El vorbeşte perfect limba româna şi întrebat asupra opiniilor sale asupra unui subiect, până la urmă le va exprima. Dar niciodată nu gândeşte (în cuvinte, imagini etc.) atunci când nu vorbeşte cu voce tare; nici nu-i 'trece nimic prin cap', cu excepţia (de bună seamă) momentului în care îşi aude propria voce şi când are senzaţiile obişnuite, la care adăugăm 'sentimentul înţelegerii' (poate are obiceiul de a-şi vorbi cu voce tare lui însuşi). Atunci când scrie la maşină o scrisoare sau când merge la magazin nu ar avea un 'flux de gânduri' intern; dar acţiunile sale sunt inteligente şi conduse de un scop, iar dacă cineva îl întreabă 'Ce faci?' este în măsură să dea un răspuns coerent.

Un asemenea personaj este întru totul imaginabil. Nimeni nu ar ezita să spună că are conştiinţă, că nu-i place *rock and roll*-ul (dacă şi-a exprimat frecvent această aversiune), chiar dacă nu are gânduri în mod conştient cu excepţia momentelor în care vorbeşte cu voce tare.

Din toate acestea urmează că: (a) nu există nici o mulţime de evenimente mentale – imagini sau altele mai abstracte – care *constituie* înţelegerea; (b) pentru a înţelege nu sunt necesare evenimente mentale. În particular, *conceptele nu pot fi identice cu obiectele mentale de nici un fel.* Să presupunem că prin obiect mental

înţelegem ceva care este accesibil introspecţiei, dar, cum am văzut, acestea pot fi absente atunci când cineva înţelege lumea (încât posedă concepte) şi, totodată, pot fi prezente atunci când nu este surprins nici un concept.

Întorcându-ne acum la critica teoriilor magice ale referinţei (problemă care l-a preocupat şi pe Wittgenstein) vedem că, pe de o parte, acele 'obiecte mentale' care sunt accesibile introspecţiei – cuvinte, imagini, simţăminte etc. – nu au referinţă în mod intrinsec mai mult decât are desenul furnicii (pentru aceleaşi motive), în timp ce tentativa de a postula obiecte mentale speciale, 'concepte', care ar avea o conexiune necesară cu referinţa şi pe care numai fenomenologii versaţi le pot pune în evidenţă, conduc la o eroare *logică* deoarece conceptele sunt (cel puţin parţial) *abilităţi* şi nu prezenţe. Doctrina că atunci când ne referim la obiectele externe trebuie să existe prezenţe mentale este nu numai o eroare a ştiinţelor naturale; ea este şi o confuzie fenomenologică şi conceptuală.

O problemă privind referinţa

De ce este surprinzător că ipoteza creierelor în container este incoerentă? Datorită înclinaţiei de a considera *ceea ce are loc în interiorul creierelor noastre* determinant pentru înţelesul şi pentru referinţa cuvintelor noastre. Dar nu este dificil de constatat că această impresie este eronată. Cuvintele indexicale precum *eu*, *aceasta*, *aici*, *acum* sunt contraexemple triviale. Eu pot avea aceeaşi stare mentală cu Henry când cred 'Eu am întârziat la slujbă' (să ne imaginăm, dacă doriţi, că Henry este geamănul meu) şi chiar într-o asemenea situaţie cuvântul 'Eu' care apare în gândul meu se referă la mine pe când cuvântul 'Eu' care apare în gândul lui Henry se referă la *Henry*.

Aş putea fi în aceeaşi stare mentală[1] atunci când gândesc 'Eu am întârziat la slujbă' marţi şi când gândesc 'Eu am întârziat la slujbă' miercuri; dar momentul la care se referă timpul verbului 'am întârziat' este diferit în cele două situaţii. Cazul termenilor naturali este mult mai subtil.

Să presupunem, revenind la exemplul din capitolul anterior ca există vorbitori de limba română pe Pământul geamăn (printr-un fel de accident miraculos ei au evoluat spre un tip de limbă română care se vorbea acum un sfert de mileniu, cu excepţia cazului pe care îl am în vedere). Presupunem că aceşti oameni nu sunt în măsură să deosebească între chimia pre- şi post- daltoniană. În particular, ei nu au noţiuni precum 'H_2O'. Să presupunem acum, că râurile şi lacurile de pe Pământul geamăn sunt umplute cu un lichid care se aseamănă cu apa, dar *nu* are compoziţia chimică H_2O. În acest caz, cuvântul 'apă', aşa cum este folosit pe Pământul geamăn, *nu* se referă la apă ci la un alt lichid (să spunem XYZ). Nu există nici o diferenţă între stările mentale ale locuitorilor celor două pământuri (să spunem în 1750) adică nu există diferenţe relevante care să dea seama de deosebirea amintită a referinţei. Referinţa

[1] Cel puţin pot fi în 'aceeaşi stare mentală' în sensul că parametrii implicaţi în procesul psihologic al gândirii mele pot avea aceleaşi valori. Starea mea mentală *globală* este diferită dacă marţi eu cred 'este marţi', pe când miercuri nu cred asta; dar o teorie care spune că sensul cuvintelor se schimbă ori de câte ori starea mentală *globală* se schimbă nu ar permite ca vreun cuvânt să aibă acelaşi sens în două ocazii diferite, aşa că s-ar ajunge la abandonarea noţiunii de sens al cuvintelor. În plus, am putea construi povestea unui Pământ geamăn în care eu şi tizul meu suntem în aceeaşi stare mentală *globală*, dar referinţa pentru 'eu' şi 'acum' rămâne diferită (calendarul de pe Pământul geamăn nu este sincronizat cu al nostru).

este diferită deoarece *lucrurile* avute în vedere sunt diferite[2]. Starea mentală, prin ea însăşi, în izolare faţă de ansamblul situaţiei, nu fixează referinţa.

Unii filosofi au adus obiecţii acestui exemplu. Aceşti gânditori sugerează că ar trebui să se spună, dacă o asemenea planetă ar fi descoperită, 'Există două tipuri de apă' şi *nu* cuvântul nostru 'apă' nu se referă la lichidul de pe Pământul geamăn. Dacă am descoperi vreodată râuri şi lacuri alcătuite dintr-un alt lichid decât H_2O care s-ar asemăna cu apa, atunci enunţul că orice apă este H_2O ar deveni fals, conform acestor critici.

Este uşor ca exemplul să fie modificat astfel încât aceste critici să fie evitate. Mai întâi, nu este nevoie ca lichidul de pe Pământul geamăn să fie *întru totul* similar cu apa. Să presupunem că avem de-a face cu un amestec din 20% alcool şi 80% apă, dar chimiştii de pe Pământul geamăn nu sunt în măsură să descopere diferenţa faţă de apă. Un asemenea lichid ar fi diferit de apă în multe privinţe; un vorbitor *tipic* ar putea să nu le constate, aşa încât să se afle în starea mentală presupusă pentru vorbitorii din 1750. De bună seamă, 'apa' de pe Pământul geamăn are alt gust pentru *noi* decât apa obişnuită; dar pentru *ei* o asemenea diferenţă nu există. De asemenea, s-ar comporta diferit la fierbere; dar trebuie ca un vorbitor de română să fi notat exact când fierbe apa pentru a asocia un standard conceptual drept conţinut al cuvântului 'apă'?

Altă obiecţie este că ar putea exista experţi pe Pământul geamăn care ştiu numeroase lucruri despre 'apă' (de pildă, că este un *amestec* a două lichide) pe care noi nu

---

[2] Vezi 'The Meaning of "Meaning"', în *Mind, Language and Reality* (*Philosophical Papers*, vol. 1), CUP, 1975 pentru o discuţie mai extinsă asupra acestui subiect.

le cunoaştem (nu erau crezute despre apă în 1750, deoarece nu era vorba de apa adevărată) astfel încât starea mentală *colectivă* a vorbitorilor de română de pe Pământul geamăn este diferită de starea mentală corespunzătoare a vorbitorilor de română din anul 1750 de pe Pământ. De aici s-ar putea trage concluzia că referinţa nu este fixată prin starea mentală individuală, ci prin starea mentală a *tuturor* membrilor unei comunităţi lingvistice.

Dar locuitorii celor două planete nu aveau o chimie suficient de dezvoltată în 1750. Dacă termenul avea acelaşi sens şi referinţă pe Pământ înainte de dezvoltarea chimiei, şi aşa s-ar întâmpla şi pe Pământul geamăn, atunci am putea merge înapoi în timp până la momentul în care stările mentale ale celor două comunităţi ar fi fost *aceleaşi* în toate aspectele relevante privitoare la apă şi să argumentăm că extensiunea era diferită *atunci* (ca şi acum) şi aşa-numitele stări mentale colective nu fixează extensiunea. Ar trebui să considerăm că referinţa se modifică odată cu dezvoltarea chimiei? Că un termen *folosit* pentru a ne referi la ambele tipuri de apă (în ciuda diferenţei de gust pentru *noi*!) are referenţi diferiţi numai datorită evoluţiei chimiei?

Dacă am considera că referinţa termenilor noştri sau ai lor se modifică odată cu chimia (după cum este capabilă să distileze lichide, să spună că apa şi alcoolul formează un amestec etc.) atunci ar trebui să acceptăm că orice descoperire ştiinţifică modifică referinţa termenilor noştri. Noi nu am *descoperi* că apa (într-un sens preştiinţific) era H_2O; mai degrabă ea ar *deveni* astfel. Pentru mine, o asemenea perspectivă este clar eronată. Ceea ce înţelegem prin apă avea aceeaşi natură de când a fost introdus termenul; şi noi am descoperit că apa în *acel înţeles* este H_2O; iar ceea ce locuitorii Pământului geamăn numesc apă a fost acelaşi lucru din mediul *lor* de când folosesc termenul

respectiv, iar experţii lor descoperă că 'apă' în *acel înţeles* este un amestec de două lichide.

Dacă suntem de acord că 'apă' nu-şi schimbă sensul (indiferent de limbaj) atunci când experţii fac o descoperire precum 'apa este H_2O' sau 'apa este un amestec de două lichide', ori nu se schimbă sensul obişnuit şi referinţa (de bună seamă s-ar putea dezvolta mai multe utilizări tehnice ca rezultat al unei asemenea descoperiri) şi că 'apa' în sensul şi referinţa obişnuită de pe Pământ nu include amestecul de alcool şi apă, atunci trebuie să admitem că ştiinţa experţilor nu dă seama de diferenţele de *sens* între cuvântul 'apă' de pe Pământ şi de pe Pământul geamăn. La fel, acestea nu dau seama de *referinţă*: deoarece am putea considera încă un Pământ geamăn unde apa ar fi un amestec *diferit*, iar cunoştinţele experţilor ar fi aceleaşi (chiar dacă insuficiente) cu acelea de pe primul Pământ geamăn. Sau, cum s-a indicat, am putea să ne imaginăm că nu există experţi pe Pământ sau pe Pământul geamăn. Chiar în asemenea situaţie, cuvântul 'apă' s-ar *referi la lucruri diferite* chiar dacă stările mentale *colective* ale celor două comunităţi ar fi aceleaşi. Ceea ce se petrece în capetele locuitorilor acelor planete nu fixează referinţa termenilor. După o expresie a lui Mill, 'substanţa în sine' completează funcţia de a fixa extensiunea termenului.

O dată ce am arătat că stările mentale (individuale sau colective) nu fixează referinţa, nu ar trebui să ne surprindă că creierele din container nu se pot referi la obiecte externe (chiar dacă ar avea aceeaşi stare mentală pe care o avem noi), aşa că nu ar spune sau gândi că ele sunt creiere în container.

Intensiuni, extensiuni şi 'lumi noţionale'

Pentru a rezolva problema modului în care termenii noştri fixează referinţa, odată ce ea nu este fixată numai prin intermediul stărilor mentale, este util să apelăm la unii termeni tehnici. În logică, lucrurile despre care este adevărat un termen alcătuiesc *extensiunea* termenului. Astfel, extensiunea termenului 'pisică' este alcătuită din mulţimea pisicilor. Dacă un termen are mai mult decât un sens, atunci cuvântului care îl exprimă i se asociază indici (aşa încât, avem de-a face cu două cuvinte în locul unuia singur), de pildă, 'iepure$_1$' – extensiune: mulţimea iepurilor – 'iepure$_2$' – extensiune: mulţimea laşilor. (Strict vorbind, în limbajul natural, extensiunea este întotdeauna, într-o oarecare măsură, fuzzy: dar vom accepta, pentru simplificare, că limitele au fost cumva fixate).

Un cuvânt precum 'Eu' care se referă la oameni diferiţi în ocazii diferite nu are o extensiune, ci o *extensiune-funcţie*: există o funcţie care determină extensiunea în fiecare context în care îl folosim. În cazul cuvântului 'Eu' extensiunea-funcţie este simplă; respectiv, este funcţia $f(x)$ ale cărei valori pentru orice vorbitor x este mulţimea care îl conţine numai pe x. Argumentul x care are ca valori tocmai elemente care provin din context (în cazul de faţă, *vorbitorul*) este considerat în semantică *indice*. Indicii sunt utilizaţi pentru momente, pentru demonstrative şi pentru alte trăsături ale contextului în tratarea lui completă sub raport semantic (dar nu vom lua în seamă acest detalii).

Mulţimea lucrurilor care alcătuiesc extensiunea cuvântului 'pisică' diferă în situaţii posibile sau 'lumi

posibile' diferite. Într-o lume posibilă *M* în care nu există pisici, extensiunea lui 'pisică' este mulţimea vidă. Dacă pisica mea Elsa ar avea urmaşi, atunci extensiunea lui 'pisică' ar avea cel puţin un membru pe care nu-l are în lumea actuală. (Putem exprima acest fapt spunând că în orice lume posibilă *M* în care Elsa a avut urmaşi, extensiunea lui 'pisică' include membri pe care nu-i are în lumea actuală).

Putem indica modul în care extensiunea unui termen variază odată cu lumea posibilă *M* exact în acelaşi fel în care am indicat cum anume extensiunea cuvântului 'Eu' se schimbă odată cu vorbitorul: printr-o *funcţie*. Presupunem o mulţime de obiecte abstracte numite 'lumi posibile' pentru a reprezenta diferitele stări de lucruri sau posibilele istorii ale lumii şi asociem termenului 'pisică' o funcţie *f*(*M*) a cărei valoare pentru fiecare lume *M* este mulţimea de obiecte posibile care sunt pisici în lumea *M*. Această funcţie, dacă îi urmăm pe Montague şi Carnap, se referă la *intensiunea*[3] cuvântului 'pisică'. În acelaşi fel, intensiunea predicatului binar 'atinge' este funcţia *f*(*M*) a cărei valoare pentru orice lume posibilă *M* este mulţimea de perechi ordonate de obiecte posibile care se ating unul cu altul în lumea *M*; intensiunea predicatului ternar '*x* este între *y* şi *z*' este funcţia a cărei valoare în orice lume posibilă constă în mulţimea tripletelor ordonate de obiecte posibile (*x*, *y*, *z*) astfel încât *x* este între *y* şi *z* etc. Intensiunea unui cuvânt precum 'eu', a cărui extensiune în orice lume este dependentă contextual, este o funcţie mai complicată având

[3] Montague R., *Formal Philosophy*, Yale University, 1974. Această utilizare a termenului 'intensiune' nu este cea tradiţională pe care o discut în 'The Meaning of "Meaning"'.

ca argumente atât lumile posibile cât şi indicii care reprezintă contextul.

Intensiunea precizează modul în care extensiunea depinde de lumea posibilă. Ea determină extensiunea în orice lume posibilă.

Motivul pentru care 'intensiunea', în acest înţeles, nu poate fi identificată cu *sensul* este că oricare doi termeni care sunt echivalenţi logic au aceeaşi extensiune în orice lume posibilă, respectiv au aceeaşi intensiune, dar o teorie care nu poate distinge între termeni cu acelaşi sens şi termeni care sunt numai echivalenţi logic este inadecvată ca teorie a *sensului*. 'Cub' şi 'poliedru cu şase feţe pătrate' sunt predicate logic echivalente. Intensiunea acestor termeni este aceeaşi, respectiv, funcţia a cărei valoare în orice lume posibilă este mulţimea cuburilor din acea lume; dar unele diferenţe de *sens* sunt pierdute dacă identificăm sensul cu funcţia amintită.

Subliniez că lumile posibile, mulţimile şi funcţiile sunt gânduri despre entităţi extra-mentale şi nu trebuie confundate cu reprezentările sau descrierile acestor entităţi.

Frege crede că sensul (Sinn) unei expresii este o entitate extra-mentală sau concept care ar putea fi cumva 'surprins' de minte. O asemenea teorie nu ne ajută în ce priveşte intensiunile în sensul acceptat de *noi*. Mai întâi, cum am notat, există o diferenţă de sens care nu este surprinsă cu ajutorul intensiunii; prin urmare, un termen nu poate consta *numai* în asocierea cu o intensiune. Mai important, dacă presupunem că nu avem un al 'şaselea simţ' care să permită accesul *direct* la entităţi extra-mentale, atunci 'surprinderea' unei intensiuni sau a vreunei entităţi extra-mentale trebuie mediată de reprezentări într-un anume fel. (Aceasta este sarcina introspecţiei.) Dar întreaga

problemă pe care o investigăm este cum anume reprezentările *pot* să ne aducă în situaţia de a ne referi la ceea ce este în afara minţii. Dacă dăm seama de noţiunea de 'surprindere' a unui *X* care este extern minţii am ajuta la rezolvarea problemei.

Dacă spun că cineva 'crede că există un pahar cu apă pe masă', atunci îi atribui capacitatea de a se referi la *apă*. Dar, cum am văzut, a fi capabil să te referi la apă necesită o legătură directă sau indirectă cu apa actuală (H_2O); judecata 'John crede că există un pahar cu apă în faţa sa' nu este numai o judecată despre ce se petrece în capul lui John, ci este în parte o judecată despre mediul lui John şi relaţia lui John cu acest mediu. Dacă John este o persoană de pe Pământul geamăn, atunci ceea ce crede John atunci când spune 'există un pahar cu apă pe masă' este că există un pahar care conţine un lichid care *în fapt* constă din apă şi alcool pe masă.

Husserl introduce un mecanism care este util atunci când vrem să vedem ce se petrece în capul cuiva fără a face presupuneri privind existenţa sau natura lucrurilor actuale la care ne referim prin gânduri: mecanismul *punerii între paranteze*[4]. Dacă punem între paranteze convingerea pe care o atribuim lui John atunci când spunem 'John crede că există un pahar cu apă pe masă', atunci ceea ce îi atribuim lui John este numai *starea mentală* a unei persoane actuale sau posibile care crede că există un pahar cu apă pe masă (în înţelesul obişnuit al punerii între paranteze). Astfel, dacă John de pe Pământul geamăn nu poate simţi diferenţa de gust dintre apă şi apă-cu-alcool, el se poate afla în aceeaşi stare mentală cu un vorbitor posibil sau actual de limba

[4] Husserl, *Ideas; General Introduction to Pure Phenomenology*, Allen and Unwin, 1969. (Originalul a apărut în 1913).

română de pe Pământ atunci când spune 'există un pahar cu apă pe masă', cu toate că atunci când se referă la apă ar avea în vedere un pahar cu whisky şi apă. Vom accepta că el are *convingerea* [există un pahar cu apă pe masă] *între paranteze.* În consecinţă, mecanismul punerii între paranteze elimină cauzele convingerilor obişnuite (toate cauzele care privesc lumea externă sau care sunt externe minţii celui care are convingerea).

Daniel Dennett a folosit recent locuţiunea 'lume noţională' legat de punerea între paranteze a lui Husserl[5]. Totalitatea convingerilor puse între paranteze ale unei fiinţe cugetătoare constituie descrierea lumii noţionale a acelei fiinţe, în înţelesul lui Dennett. Astfel, locuitorii Pământului geamăn au, în linii mari, aceeaşi lume *noţională* (chiar aceeaşi apă *noţională*) pe care o avem noi deşi trăiesc pe o planetă în realitate *diferită* (ne referim la lucrurile care sunt diferite în mod actual, cum este apa). La fel, creierele din container, amintite în capitolul anterior, ar putea avea aceeaşi lume noţională pe care o avem noi până în cel mai mic detaliu, dacă doriţi; numai că nici unul dintre termenii lor nu are referinţă în lumea externă. Teoria tradiţională a sensului presupune că lumea noţională a unei fiinţe cugetătoare determină intensiunile termenilor săi (şi acestea, împreună cu faptul că o lume posibilă particulară *M,* cum este cea actuală, determină extensiunile termenilor şi valorile de adevăr ale tuturor enunţurilor). Am văzut că teoria tradiţională a sensului este eronată; acesta este motivul pentru care literatura contemporană conţine numeroase concepte diferite (de pildă, 'intensiune', 'lume noţională') dar nici un concept

[5] Dennett D., 'Beyond Belief', *Thought and Object*, Andrew Woodfield (ed.), OUP.

unitar pentru *sens*. 'Sensul' rămâne *neclar*. Avem doar impresia clarificării. Dacă intensiunea şi extensiunea nu sunt fixate prin lumea noţională, atunci *cum sunt ele fixate?*

Perspectiva tradiţională asupra interpretării

Cea mai obişnuită manieră de a interpreta limbajul, care este fixată de către noi în mod colectiv dacă nu individual, este asociată cu noţiunile de *constrângere operaţională* şi *constrângere teoretică*. Prima formă de constrângere este concepută mai degrabă naiv; noi *stipulăm* pur şi simplu (în mod convenţional) că o anumită propoziţie (să apunem 'Prin cablu trece electricitate') este adevărată dacă şi numai dacă se constată un anume rezultat în urma unui test (acul unui voltmetru deviază de la poziţia de echilibru, sau, în versiune fenomenologică, *impresia vizuală pe care o am privind acul voltmetrului oscilând*). Acest tip de operaţionalism brut nu mai are aderenţi deoarece se apreciază că: (1) legăturile dintre teorie şi experienţă sunt probabiliste şi nu pot fi formalizate corect drept corelaţii perfecte (chiar dacă există un curent care trece prin cablu există întotdeauna evenimente cu o probabilitate scăzută sau condiţii de mediu care ar putea împiedica acul voltmetrului să devieze) şi (2) legăturile probabiliste nu sunt corelaţii *semantice*, ci depind de teoria empirică, iar aceasta este subiect de revizuire. Pentru operaţionalistul naiv, de fiecare dată când este descoperită o nouă metodă pentru a testa dacă ceva este din aur, sensul cuvântului 'aur' se modifică (nu am putea spune că a fost *descoperit* un nou test pentru aur). Conform operaţionalismului, teoriile sunt

testate *propoziţie cu propoziţie* (sensul stipulat operaţional pentru propoziţiile individuale ne spune cum anume trebuie să procedăm pentru a testa teoria); conform unei concepţii mai noi, 'o teorie se confruntă cu testele ca un întreg', după cum argumentează Quine.

Operaţionalismul poate fi nuanţat aşa încât obiecţiile de mai sus să fie evitate. Astfel, clasa interpretărilor poate fi restrânsă (asocierea de intensiuni pentru predicatele din limbaj) ajungându-se la constrângeri de forma: 'o interpretare este admisibilă atunci când, *de cele mai multe ori*, propoziţia *S* este adevărată când condiţia empirică *E* este îndeplinită' (respectiv, 'atunci când, *de cele mai multe ori*, propoziţia *S* este *falsă* când condiţia de experienţă *E* este îndeplinită'). Un asemenea model pentru constrângere corespunde relaţiei *probabiliste* între valoarea de adevăr a propoziţiilor şi experienţă. În al doilea rând, asemenea reguli ar fi adaptabile pentru teoriile în schimbare. Mai degrabă, decât să fie introduse stipulaţii cum sunt cele are raţionalismului brut, s-ar putea apela la reguli nuanţate; împreună cu Peirce (care a scris cu 50 de ani înainte ca Bridgman să vorbească de operaţionalism!) am putea considera că mulţimea ideală de constrângeri operaţionale se modifică pentru a aproxima evoluţia cercetării empirice şi nu este stipulată. Pe scurt, constrângerile operaţionale ar putea fi considerate impuse de cercetarea raţională, de ceea ce este observat, experimentat sau raţionat, constrângerile ar fi un fel de 'echilibru raţional' care conduce la interpretări pentru termenii noştri; constrângerile sunt într-adevăr acceptate la un moment dat cu un statut de aproximare raţională.

Un asemenea punct de vedere este compatibil cu argumentarea lui Quine că legătura între experienţă şi teorie este o problemă ce trebuie supusă revizuirii ca orice

alt aspect din corpusul integrat al cunoaşterii. Aceste revizuiri nu trebuie văzute de fiecare dată ca 'schimbări de sens': ele pot să fie adesea simple eforturi de a preciza mai bine despre ce este vorba, prin care vechile constrângeri teoretico-operaţionale apar ca neadecvate.

Pe lângă restrângerea clasei interpretărilor admisibile prin intermediul constrângerilor operaţionale (sau a aproximărilor succesive spre mulţimea ideală de constrângeri invocată de Peirce), am putem avea în vedere şi constrângeri care se referă la proprietăţile formale ale teoriilor. De pildă, 'o interpretare admisibilă este aceea în care devine adevărat că efecte diferite au întotdeauna cauze diferite'. Kant a susţinut că asemenea 'constrângeri teoretice' sunt părţi ale raţionalităţii însăşi: noi *impunem* principiul determinismului în lume, mai degrabă decât îl descoperim. Sub această formă, constrângerile sunt prea puternice: *preţul* păstrării determinismului ar putea fi o prea mare complexitate a sistemului nostru de cunoştinţe în ansamblu. Totuşi acest tip de constrângere poate fi nuanţat la fel cum au fost constrângerile operaţionale. (De pildă, s-ar putea cere ca determinismul să fie păstrat când 'costul' complicării altor părţi ale teoriei nu este prea mare; în această formă, consider că o asemenea constrângere poate fi acceptată).

Constrângerile teoretice sunt considerate adesea constrângeri ale interpretării teoriei, dar ele pot fi regândite cu uşurinţă încât să joace rolul menţionat aici. Astfel, dacă un autor stabileşte constrângerea de 'conservare' ca o constrângere de acceptare teoretică ('nu accepta o teorie care cere să se renunţe la multe dintre convingerile deja acceptate dacă există o teorie – la fel de simplă – care conservă acele convingeri şi este în acord cu observaţia'), atunci o putem reformula ca o constrângere de interpretare

astfel: 'o interpretare admisibilă dă seama de propoziţiile adevărate acceptate de-a lungul timpului cu excepţia celor care conduc la complicaţii excesive în cadrul teoriei alcătuită dintr-o mulţime de propoziţii adevărate sub interpretarea respectivă, sau generează schimbări prea mari în cadrul constrângerilor operaţionale'. Din nou, este prezentă teza că nu este posibilă nici o logică inductivă decât dacă se impune o ordine *a priori* (numită 'ordine de simplitate' sau 'ordine de plauzibilitate') sau ipoteze care pot fi acceptate dat fiind anumite evidenţe empirice particulare (deşi ordonarea poate ea însăşi să fie diferită în raport cu contextele experimentale sau observaţionale); constrângerea că 'mulţimea propoziţiilor care sunt adevărate sub o interpretare admisibilă nu trebuie să fie inferioară în ordinea simplităţii faţă de orice altă mulţime care are aceleaşi consecinţe observaţionale sau experimentale' corespunde constrângerii din logica inductivă de a accepta ipoteza cea mai simplă (sau mai 'plauzibilă') compatibilă cu observaţiile.

În literatura de filosofie a ştiinţei au fost propuse şi alte tipuri de constrângeri teoretice. Există constrângeri care, precum 'simplitatea', se referă la proprietăţile propoziţiilor acceptabile şi constrângeri care se referă la *istoria* cercetării prin care sunt acceptate propoziţiile. Detaliile nu sunt relevante aici. Motivele pentru care acordăm atenţie ideii că interpretările admisibile ale limbajului (în sensul de atribuiri-de-intensiune pentru termenii limbajului) sunt determinate de constrângeri operaţionale sau teoretice sunt imediate: faptul că o persoană are sau nu o experienţă de un anume fel este o problemă de judecată mentală (problemele filosofice ale 'experienţei' nu sunt luate în considerare). Astfel, dacă o teorie implică sau selectează o propoziţie care este asociată

cu o experienţă *E* printr-o constrângere operaţională de un anume tip, probabilistic sau de altă natură, atunci fiinţa cugetătoare poate *cunoaşte* dacă teoria funcţionează, sau dacă există ceva problematic aici, cel puţin în acest caz, după cum experienţa *E* are loc sau nu. Deoarece constrângerile pe care le folosim pentru a testa teoriile fixează *şi ele* extensiuni ale termenilor, estimarea asupra 'modului de funcţionare' a teoriei este în acelaşi timp o estimare a *adevărului* său. Cunoştinţele vorbitorului despre aceste constrângeri sunt cunoştinţe asupra intensiunii termenilor, aşa încât dobândirea unei semantici corecte ne-ar spune, pentru orice teorie asupra lumii, cum ar fi lumea, dacă teoria ar fi adevărată.

Afară de aceasta, dacă idealizăm prin ipoteza unor fiinţe cugetătoare care posedă una despre alta ceea ce economiştii numesc 'informaţie perfectă', fiecare asemenea fiinţă cunoaşte structura formală a teoriei acceptate *T* şi istoria programului de cercetare căreia îi aparţine, convingerile anterioare despre ce se conservă etc. În acest fel, fiecare fiinţă cugetătoare este în măsură să *cunoască* dacă au loc sau nu constrângerile teoretice. (Dacă nu idealizăm prin presupunerea informaţiei perfecte, atunci încă putem spune despre ansamblul fiinţelor cugetătoare că ar putea cunoaşte aceste fapte).

Pe scurt, dacă perspectiva de aici este corectă, atunci am avea o abordare elegantă despre modul în care sunt fixate intensiunile şi extensiunile (în principiu – deoarece detaliile sunt prea complicate pentru a fi puse în evidenţă în acest stadiu al cercetărilor). Cu toate acestea, din păcate, perspectiva tradiţională nu funcţionează!

De ce nu funcţionează perspectiva tradiţională asupra interpretării

Dificultatea de care se izbeşte perspectiva tradiţională este că încearcă să fixeze intensiunile şi extensiunile termenilor individuali prin fixarea condiţiilor de adevăr ale întregii propoziţii. După cum am arătat, teza privind constrângerile operaţionale şi cele teoretice (cercetările raţionale ar accepta un fel de limită ideală a cercetării) determină care propoziţii ale limbajului sunt *adevărate*. Chiar dacă aşa stau lucrurile, constrângerile nu pot determina la ce se *referă* termenii noştri. Nu există constrângeri prin care un asemenea obiectiv să fie atins direct. Iar strategia indirectă formulând constrângerile care circumscriu mulţimea propoziţiilor adevărate, şi apoi sperând că prin determinarea valorilor de adevăr ale propoziţiilor ca întregi am putea fixa cumva la ce se referă *termenii* care apar în aceste propoziţii, nu funcţionează, aşa cum a arătat Quine[6]. Eu voi extinde rezultatele 'indeterminării' precedente. Voi argumenta că, chiar dacă avem constrângeri de orice natură care determină valorile de adevăr într-un limbaj *în raport cu orice lume posibilă*, referinţa termenilor individuali rămâne nedeterminată. De fapt, este posibil să interpretăm limbajul în întregul său în moduri cu totul diferite, compatibile cu cerinţele de adevăr ale fiecărei propoziţii în orice lume posibilă. Pe scurt, nu numai că perspectiva tradiţională nu funcţionează; *nici un punct de vedere care se rezumă la fixarea valorilor de adevăr ale propoziţiilor în întregul lor nu poate fixa referinţa*, chiar dacă valorile de adevăr respective au loc în orice lume posibilă.

---

[6] Vezi lucrarea sa 'Ontological Relativity', *Ontological Relativity and Other Essays*, Columbia U.P., 1969.

Demonstraţia în detaliu este pur tehnică şi cred că este mai nimerit să o prezint într-o *Anexă*. Aici doar voi ilustra metoda folosită în cadrul demonstraţiei.

Să considerăm propoziţia:

(1) O pisică este pe un covor. (Aici şi în cele ce urmează 'este pe' nu are *timp*, putând să însemne 'este, a fost, va fi pe').

Sub interpretarea standard, această propoziţie este adevărată în acele lumi posibile în care există cel puţin o pisică pe cel puţin un covor la un moment oarecare (trecut, prezent sau viitor). În plus, 'pisică' se referă la toate pisicile, iar 'covor' la toate covoarele. Voi arăta că propoziţia (1) poate fi reinterpretată astfel încât în lumea *actuală* 'pisică' să se refere la *cireşe* iar 'covor' la *mese* fără a afecta valoarea de adevăr a propoziţiei (1) în orice lume posibilă. ('este pe' rămâne cu aceeaşi interpretare).

Propoziţia (1) va primi o nouă interpretare în care apar noile sensuri:

(a) O pisică* este pe un covor*.

Definiţia proprietăţii a fi o pisică* (respectiv, un covor*) este dată de următoarele trei situaţii:

(a) Unele pisici sunt pe unele covoare şi unele cireşe sunt pe unele mese.

(b) Unele pisici sunt pe unele covoare şi nici o cireaşă nu este pe nici o masă.

(c) Nici una dintre situaţiile anterioare.

Iată definiţia celor două proprietăţi:

DEFINIŢIA PENTRU 'PISICĂ*'
x este o pisică* dacă şi numai dacă are loc situaţia (a) şi *x* este o cireaşă; sau are loc situaţia (b) şi *x* este o pisică; sau se petrece situaţia (c) şi *x* este o cireaşă.

DEFINIŢIA PENTRU 'COVOR*'
x este un covor* dacă şi numai dacă are loc situaţia (a) şi *x* este o masă; sau are loc situaţia (b) şi *x* este un covor; sau se petrece situaţia (c) şi *x* este un quark.

Acum, în lumea posibilă unde are loc (a), 'O pisică este pe un covor' este adevărată, şi 'O pisică* este pe un covor*' este, de asemenea, adevărată (deoarece o cireaşă se află pe o masă şi toate cireşele sunt pisici* şi toate mesele sunt covoare* în lumile de acest tip). Dacă în lumea actuală unele cireşe sunt pe unele mese, lumea actuală este de tipul descris mai sus şi în lumea actuală 'pisică*' se referă la cireşe, iar 'covor*' se referă la mese.

În lumile posibile unde are loc (b), 'O pisică este pe un covor' este adevărată şi 'O pisică* este pe un covor*' este de asemenea adevărată (deoarece în lumile în care are loc (b), 'pisică' şi 'pisică*' sunt termeni coextensivi, la fel cu 'covor' şi 'covor*'). (Să remarcăm că, deşi pisicile sunt pisici* în unele lumi – cele în care are loc (b) – ele nu sunt pisici* în lumea efectivă).

În lumile posibile care cad sub cazul (c), 'O pisică este pe un covor' este falsă, şi 'O pisică* este pe un covor*' este, de asemenea, falsă (deoarece o cireaşă nu poate fi un *quark*).

Pe scurt, am văzut că *în orice lume posibilă* o pisică este pe un covor dacă şi numai dacă o pisică* este pe un covor*. Astfel, dacă reinterpretăm cuvântul 'pisică' prin atribuirea unei intensiuni, este interpretat şi pisică*, la fel şi pentru 'covor'. Aceasta are ca efect că 'O pisică este pe un covor' înseamnă 'O pisică* este pe un covor*' după cum a fost definit; iar această situaţie conduce la aceleaşi valori de adevăr în orice lume posibilă.

În *Anexă* arăt că o reinterpretare mai complicată de acest tip poate fi realizată pentru toate propoziţiile unui limbaj. Urmează că întotdeauna există infinit de multe interpretări diferite pentru predicatele unui limbaj pentru aceleaşi valori de adevăr 'corect' atribuite pentru propoziţii în toate lumile posibile, *fără a avea importanţă dacă aceste valori de adevăr 'corecte' sunt selectate.* Quine argumentează pentru o concluzie similară în *Cuvânt şi obiect*; în exemplul lui Quine (aplicat la română) 'Acolo se află un iepure' era interpretat prin 'Acolo se află o secvenţă de iepure' (unde 'o secvenţă de iepure' este o secţiune spaţială tridimensională a iepurelui patru-dimensional din spaţiu-timp) sau, ceea ce este similar, 'Iată un exemplar al iepurimii'. (Ultima variantă reface inclusiv forma sintactică a propoziţiei sau cel puţin gramatica sa logică). Quine arată aici ceea ce am remarcat deja, *condiţiile de adevăr pentru întreaga propoziţie* nu determină referinţa. Deoarece 'secvenţa de iepure' şi 'iepurimea' ca şi 'părţile nedetaşate de iepure', sunt strâns legate de iepuri, s-ar putea merge mai departe decât a făcut-o Quine în *Cuvânt şi obiect* cu impresia că toate reinterpretările care lasă valoarea de adevăr a unei propoziţii neschimbată sunt strâns legate de interpretarea standard (în modul în care părţile-de-iepure şi iepurimea sunt legate de iepuri). Argumentul dezvoltat în *Anexă* şi ilustrat în acest capitol arată că adevărul pentru 'O pisică este pe un covor' nu exclude ca 'pisică' să se refere la *cireşe.*

'Intrinsec' şi 'extrinsec'

Poate că primul lucru la care ne gândim atunci când suntem confruntaţi cu interpretări non-standard, cum ar fi aceea prin care 'pisică' devine *pisică** şi 'covor' devine *covor** este respingerea şi considerarea lor ca un paradox neimportant. Dar paradoxurile autentice nu sunt niciodată neimportante, ci arată întotdeauna că există ceva în neregulă cu modul în care obişnuim să gândim. O altă reacţie ar putea fi protestul că *pisică** şi *covor** sunt proprietăţi 'ciudate'; cu siguranţă termenilor noştri le corespund proprietăţi sensibile (precum *a fi pisică* sau *a fi covor*) şi nu asemenea proprietăţi 'caraghioase'. S-ar putea explica modul în care *pisică** (sau mai curând *pisicitatea** ori *pisica*-tatea*) este o proprietate caraghioasă arătând că cineva ar putea 'construi o maşină' pentru a 'inspecta' lucrurile şi a 'spune' dacă există sau nu pisici (o fiinţă umană este o asemenea 'maşină'), dar nimeni nu poate construi o maşină care va spune (în orice lume asemănătoare cu a noastră privind legile şi condiţiile generale) dacă ceva este *pisică**. Dacă maşina sau persoana priveşte la ceva şi vede că nu este nici pisică nici cireaşă, atunci spune că *nu* este o *pisică**; dar dacă acel lucru este fie pisică fie cireaşă, atunci maşina sau persoana are nevoie să fie informată asupra adevărului pentru 'O pisică este pe covor' sau 'O cireaşă este pe masă' pentru a stabili dacă lucrul inspectat sau văzut este *pisică**, aşa încât valorile de adevăr se află dincolo de ceea ce putem învăţa prin examinarea obiectului prezentat.

Din păcate, 'vede' ar putea fi reinterpretat (de pildă, ca vede*) aşa încât cele două propoziţii

(3) John (sau altcineva) vede o pisică

şi

(4) John vede* o pisică*,

să aibă aceeaşi valoare de adevăr în orice lume posibilă (prin metoda care este descrisă în *Anexă*). Astfel, orice persoană care ar vedea o pisică, ar vedea* o pisică*; experienţa pe care o avem atunci când vedem o pisică *este* experienţa avută atunci când vedem* o pisică* etc. În mod similar, putem reinterpreta 'inspectează' şi 'spune' astfel încât, atunci când o maşină inspectează o pisică, ea inspectează* o pisică* şi când 'spune' despre ceva că este pisică, spune* că este o pisică*.

Pentru a da un exemplu (sugerat de Nozick) să presupunem că jumătate dintre noi (bunăoară femeile) folosind 'pisică' înţeleg 'pisică*', prin 'covor' ele înţeleg 'covor*', 'a privi' este înţeles 'a privi*', 'a spune' are înţelesul, pentru ele, 'a spune*' etc. Să presupunem că cealaltă jumătate (bărbaţii) folosesc 'pisică' pentru a denota pisici, 'covor' pentru a denota covoare, 'a privi' pentru a denota privirea etc. Cum am putea să punem în evidenţă asemenea diferenţe?[7] (Dacă este întrebat un bărbat la ce se referă 'pisică', el va răspunde 'la pisici, de bună seamă', ceea ce va răspunde şi o femeie.

Urmarea este că, un cercetător care *construieşte o maşină pentru a inspecta lucrurile şi pentru a spune dacă acestea sunt pisici* deosebeşte pisicile de pisici* în cazul în care 'inspectează' şi 'spune' se referă la inspectare şi spunere, aşa că nu este mai uşor să spunem cum este fixată referinţa *acestor* cuvinte decât referinţa cuvântului 'pisică'.

---

[7] O femeie ar putea răspunde că supoziţia că ea se referă la pisici* când spune 'pisică' este *incoerentă* (deoarece în *interiorul* limbajului ei ori de câte ori se referă la o 'pisică', aceasta *este* o pisică). Acest răspuns nu este de mare ajutor; nu exclude posibilitatea să numească *pisică* ceea ce bărbaţii numesc *pisică** şi reciproc; aceasta este poziţia lui Nozick.

S-ar putea ca, atunci când privesc spre ceva şi gândesc că este o pisică, 'reprezentările mentale' pe care le am, imaginile vizuale şi cele tactile, verbalizarea gândului 'pisică' etc. se *referă* la pisicitate şi la diferite alte proprietăţi fizice sau biologice (a avea o anumită formă, a avea o anumită culoare, a aparţine unei anume specii etc.) şi nu la contrarele acestora; aceasta ar putea să fie adevărat, dar chiar aşa referinţa este fixată pe căi diverse. Asta este ceea ce am dorit să explicăm şi nu explicaţia căutată.

'Dar, s-ar putea obiecta, *definiţiile* pentru 'pisică*' şi pentru 'covor*' date anterior se referă la alte lucruri (cireşe pe mese şi pisici pe covoare) aşa încât semnifică în mod *extrinsec* proprietăţi ale obiectelor care au asemenea proprietăţi. În lumea actuală, orice cireaşă este o pisică*, dar nu ar fi o pisică*, chiar dacă proprietăţile intrinseci ar fi aceleaşi, dacă nu s-ar afla pe o masă. În contrast, faptul că ceva este pisică depinde numai de proprietăţile intrinseci.' Oare distincţia între proprietăţi *intrinseci* şi *extrinseci* ne permite să formulăm reguli şi să determinăm interpretările 'dubioase'?

Impedimentul care afectează această sugestie constă într-o anumită simetrie a relaţiei dintre 'pisică' şi 'covor', respectiv, 'pisică*' şi 'covor*'. Să presupunem că definim 'cireaşă*' şi 'masă*' aşa încât în lumile posibile în care are loc (a) cireşele* sunt pisici şi mesele* sunt covoare; în lumile de sub cazul (b) cireşele* sunt cireşe iar mesele* sunt mese; iar în lumile posibile de tipul (c) cireşele* sunt pisici şi mesele* sunt fotoni. În acest caz putem defini 'pisică' şi 'covor' prin sensul *-termenilor corespunzători, astfel:

(a*) Unele pisici* sunt pe unele covoare* şi unele cireşe* sunt pe unele mese*.

(b*) Unele pisici* sunt pe unele covoare* şi nici o cireaşă* nu este pe vreo masă*.

(c*) Nu are loc nici un caz de mai sus.

Straniu, dar aceste situaţii sunt tocmai precedentele (a), (b) şi (c) sub o nouă descriere. Acum să definim:

DEFINIŢIA PENTRU 'PISICĂ'

x este o pisică dacă situaţia (a*) are loc şi *x* este o cireaşă*; sau situaţia (b*) are loc şi *x* este o pisică*; sau (c*) are loc şi *x* este o cireaşă*. (Să remarcăm că în toate trei cazurile pisicile ajung să fie *pisici*).

DEFINIŢIA PENTRU 'COVOR'

x este un covor dacă şi numai dacă are loc situaţia (a*) şi *x* este o masă*; sau în cazul (b*) şi *x* este un covor*; sau în cazul (c*) şi *x* este un quark*. (Presupunem că quark* este definit astfel încât în cazul (c*) quark*-urile sunt *covoare* iar în toate trei cazurile covoarele ajung să fie *covoare*).

Prin urmare, într-un limbaj unde 'pisică*', 'covor*' etc. sunt luate ca proprietăţi primitive, 'pisică' şi 'covor' se referă la proprietăţi 'extrinseci', proprietăţi în a căror definiţie sunt menţionate alte obiecte decât *x*; în vreme ce, relativ la un limbaj normal, în care 'pisică' şi 'covor' se referă la pisicitate şi covoreitate (ştii la ce proprietăţi mă refer, dragă cititorule!), 'pisică*' şi 'covor*' se referă la proprietăţi extrinseci. Altfel spus, a fi 'intrinsec' sau

'extrinsec' este relativ la alegerea proprietăţilor considerate *de bază*; nici o proprietate nu este intrinsecă sau extrinsecă prin ea însăşi.

'Supravieţuire' şi evoluţie

În zilele noastre se consideră că evoluţia a produs într-un fel sau altul o corespondenţă între cuvintele sau reprezentările noastre mentale şi lucrurile externe; unii spun că nu am fi *supravieţuit* dacă nu ar fi avut loc o asemenea corespondenţă şi că această corespondenţă este, cel puţin într-o formă primitivă, relaţia de referinţă.

Dar ce au de-a face 'corespondenţa' şi 'referinţa' cu *supravieţuirea*? Cum anume *adevărul* are importanţă pentru supravieţuire?

Asupra acestor chestiuni opiniile sunt împărţite. Unii filosofi cred că nu am fi supravieţuit dacă convingerile noastre (suficient de multe dintre ele) nu ar fi fost *adevărate*. Alţi filosofi afirmă că şi cele mai bine stabilite convingeri, prin mijloace ştiinţifice, nu sunt adevărate sau cel puţin că nu avem motive să credem acest lucru. Thomas Kuhn a sugerat că convingerile noastre 'se referă' la obiecte în *interiorul* acestor convingeri (cumva în felul în care 'Hamlet' se referă numai la o persoană din piesă); succesul ştiinţei este explicat prin încercare-şi-eroare şi nu prin corespondenţă cu lucrurile reale, spune Kuhn. Bas van Fraassen, într-o carte recentă, argumentează că succesul teoriei nu are nevoie de adevăr, ci de 'adecvare empirică', bunăoară de predicţii a căror corectitudine este stabilită prin observaţii. El explică succesul (sau 'adecvarea empirică') ştiinţei ca produs al mecanismului încercare-şi-eroare.

Dacă aceşti filosofi au dreptate, atunci întregul demers de a invoca evoluţia ca justificare pentru convingerea într-o relaţie *obiectivă* a referirii cade. Evoluţia, într-o perspectivă de o asemenea factură instrumentalistă, are numai rolul de a stabili o corespondenţă între unii termeni (termenii observaţionali) şi 'posibilităţile permanente de senzaţii'. O asemenea corespondenţă nu este *referire*, decât în cazul în care am abandona ideea că lucrurile externe (cele observabile) sunt mai mult decât constructe sau posibilităţi de senzaţii.

Eu consider că filosofii ceilalţi au dreptate (cei care spun că nu am fi supravieţuit dacă suficient de multe convingeri ale noastre nu ar fi fost *adevărate*).

Motivul pentru care cred acest lucru este că mecanismul încercare-şi-eroare nu explică *de ce* teoriile noastre sunt 'adecvate empiric'; *adecvarea empirică* poate fi explicată numai prin referire la *interacţiunile* dintre om şi mediu care explică de ce *încercarea-şi-eroarea are succes*. (Încercarea-şi-eroarea nu are succes în toate întreprinderile noastre, la urma urmei!) Ipoteza că interacţiunea produce în mintea noastră teorii *false* care din *întâmplare* generează predicţii care au succes înseamnă a admite o serie întreagă de *coincidenţe* inexplicabile. Dar cum anume, faptul că convingerile noastre sunt adevărate (cu aproximaţie) explică supravieţuirea?

Unele dintre convingerile noastre sunt strâns legate de *acţiune*. Dacă eu cred propoziţia 'Voi primi ceva foarte valoros dacă apăs acest buton' (presupun că înţeleg această propoziţie într-un mod normal sau, cel puţin, apreciez corect ceea ce este pus între paranteze sau ceea ce este 'noţional' cu privire la această convingere) atunci voi duce mâna şi voi apăsa butonul. Să numim convingerile de forma 'Dacă fac *x* atunci obţin...' unde locul gol este pentru un *scop*, *convingeri*,

directive. Dacă prea multe dintre aceste convingeri directive sunt false, voi săvârşi prea multe acţiuni fără succes; în acest fel adevărul convingerilor directive (în număr suficient) este necesar pentru supravieţuire.

Acum, convingerile directive sunt derivate din multe alte credinţe: convingerile despre caracteristicile şi puterile care acţionează asupra lucrurilor externe şi convingerile despre propriile noastre caracteristici şi puteri. Dacă aceste convingeri ar fi cu prioritate false, n-ar fi o simplă coincidenţă faptul că ele conduc la predicţii adevărate privind experienţa şi la convingeri directive adevărate? Astfel, dacă suficient de multe dintre convingerile noastre directive sunt adevărate, şi *cea mai bună explicaţie* a acestui fapt este că multe dintre celelalte convingeri ale noastre (cele care constituie 'teoria comună asupra lumii') sunt cel puţin aproximativ adevărate, suntem justificaţi în convingerea că teoria comună asupra lumii este cel puţin cu aproximaţie adevărată şi că nu am fi supravieţuit dacă nu s-ar întâmpla aşa.

Să ne imaginăm acum că unii dintre noi se referă în mod efectiv la lucruri care sunt asignate de către termenii noştri prin interpretarea non-standard *J* (descrisă în *Anexă*). Această interpretare este la fel cu interpretarea standard pentru termenii care se referă la lumea noastră noţională, la senzaţiile noastre, la dorinţele noastre etc. Astfel, propoziţia 'Mi se pare că apăs butonul' când o înţelegem prin operaţia 'punerii între paranteze' (cu sensul că am o anumită experienţă a apăsării voite a butonului) nu are numai aceleaşi condiţii de adevăr dar şi aceeaşi *interpretare* sub *J* precum sub interpretarea 'normală' *I*; la fel se întâmplă pentru 'Mi se pare că dobândesc satisfacţia pe care o aşteptam'.

Acum, dacă suficient de multe dintre convingerile noastre directive sunt adevărate sub interpretarea non-standard *J*, atunci cu certitudine vom avea succes şi vom *supravieţui* (deoarece dacă nu am fi în viaţă nu am atinge scopurile propuse) şi vom avea urmaşi (deoarece dacă ei nu ar trăi atunci nu ar atinge scopurile respective). Pe scurt, *J*-adevărul a (suficient de multe) convingeri directive este la fel de bun pentru a explica 'succesul evoluţiei' ca şi *I*-adevărul. În fapt, acesta este *I*-adevărul când condiţiile de adevăr pentru *fiecare* propoziţie (nu numai convingerile directive) sunt aceleaşi sub *J* ca şi sub *I*. Convingerile mele directive nu numai că sunt asociate cu aceeaşi *experienţă* subiectivă sub interpretarea *I* şi sub interpretarea *J*; ele au aceleaşi *condiţii de adevăr*. Din punctul de vedere al 'evoluţiei', este nevoie ca suficient de multe convingeri să fie adevărate sub *orice* interpretare care conectează aceste convingeri cu *acţiunile* relevante. Evoluţia poate produce în mine o tendinţă de a avea convingeri *adevărate* (de un anumit tip); dar aceasta înseamnă că evoluţia afectează prin mediere lingvistică sau conceptuală supravieţuirea *pe calea* tendinţei sale de a produce în noi sisteme de reprezentare a căror propoziţii sau 'interpretări ale propoziţiilor' au anumite *condiţii de adevăr* (şi anumite condiţii de *acţiune*, sau 'reguli de ieşire din limbaj'). *Dar condiţiile-de-adevăr pentru toate propoziţiile, aşa cum am arătat, nu determină referinţa părţilor propoziţiei* (nici adăugarea 'regulilor de ieşire din limbaj' nu ne ajută, deoarece acestea sunt aceleaşi şi sub *J*). Urmează că este o eroare să se creadă că evoluţia determină o corespondenţă *unică* (sau măcar o corespondenţă suficient de precisă) între expresiile prin care realizăm referirea şi lucrurile externe.

Intenţii: pur şi impur

Am văzut că natura nu selectează vreo corespondenţă unică între termenii noştri şi lucrurile externe. Natura lasă pe seama noastră procesarea cuvintelor şi a semnelor gândite astfel încât suficient de multe dintre convingerile noastre directive să fie adevărate şi suficient de multe dintre acţiuni să contribuie la 'bunăstarea noastră genetică'; dar acest mecanism lasă referirea nedeterminată. W.V. Quine a arătat cum este referinţa de fapt – nedeterminată! El crede că este o simplă iluzie că termenii limbajului nostru au contrapărţi bine determinate. După cum el însuşi afirmă:

> De aceea, să ne reîntoarcem la notarea standard, cu o listă de predicate interpretate şi un domeniu de valori pentru variabile. Propoziţiile adevărate dintr-un asemenea limbaj rămân adevărate sub nenumărate reinterpretări ale predicatelor şi revizuiri ale domeniului de valori. Într-adevăr, orice domeniu de aceeaşi mărime poate fi folosit pentru a obţine interpretări convenabile ale predicatelor. Dacă domeniul de valori este infinit, orice domeniu infinit conduce la rezultate similare; aceasta este teorema Löwenheim-Skolem. Propoziţiile adevărate rămân adevărate indiferent de asemenea schimbări.
>
> Poate că atunci, prima noastră grijă trebuie să fie cu privire la adevărul propoziţiilor şi la condiţiile lor de adevăr, mai degrabă decât la referinţa termenilor.

În capitolul următor voi explora alternativa sugerată aici pornind de la ideea care este premisa întregii discuţii: cuvintele stau într-un fel de corespondenţă biunivocă cu lucrurile (independent de discurs) sau cu mulţimile de lucruri. Mi se pare, însă, că există o modalitate mai simplă: de ce să nu spunem că *intenţiile* noastre fixează referinţa termenilor, în mod implicit sau explicit?

La începutul discuţiei din capitolul precedent, am respins o asemenea variantă ca neavând conţinut informativ pe motivul că a avea intenţii (într-un fel relevant) presupune abilitatea de a ne referi. Este momentul să dezvolt această remarcă.

Problema este că noţiunile de 'intenţie' şi 'stare mentală' sunt ambigue. Să numim o stare mentală *pură*, dacă prezenţa sau absenţa ei depinde numai de ceea ce se petrece în 'interiorul' vorbitorului. De pildă, dacă am sau nu o durere depinde numai de ceea ce se petrece în interiorul meu, dar cunoştinţa că zăpada este albă depinde atât de ceva care se petrece în interiorul meu (a avea convingerea că zăpada este albă) cât şi de faptul dacă zăpada este sau nu albă, adică de ceva ce se află în afara minţii şi corpului meu. Astfel, *durerea* este o stare mentală pură, pe când *cunoaşterea* este o stare mentală *impură*. Există o componentă (pură) mentală a cunoaşterii dar există şi o componentă non-mentală corespunzătoare condiţiilor în care convingerile unui om reprezintă cunoaştere numai dacă sunt *adevărate*. Nu sunt într-o stare de cunoaştere că zăpada este albă dacă nu sunt într-o stare mentală pură corespunzătoare; dar a fi într-o stare mentală pură corespunzătoare nu este niciodată *suficient* pentru a cunoaşte că zăpada este albă; la aceasta trebuie să coopereze şi lumea.

Ce este o convingere? Am definit convingerea *pusă între paranteze* ('lumea noţională') astfel încât atunci când avem o convingere pusă între paranteze precum [pe masă se află un pahar cu apă] sau o lume noţională care include existenţa paharului pe masă ne aflăm într-o stare mentală pură. Dar, în acord cu cele arătate anterior, *convingerea că există un pahar cu apă pe masă* (fără vreo 'punere în paranteze') presupune că, prin cuvântul 'apă' ne referim *în*

mod efectiv la apă iar aceasta depinde de natura anumitor 'paradigme' şi de relaţia cauzală directă sau indirectă cu aceste paradigme etc. Când eu am convingerea de mai sus, corespondentul meu de pe Pământul geamăn are o convingere între paranteze asemănătoare dar nu aceeaşi convingere , deoarece cuvântul său 'apă' se referă la apă-cu-alcool şi nu la apă. Pe scurt, convingerea că 'pe masă se află un pahar cu apă' este o stare mentală *impură*. (Creierele din container nu ar putea fi în această stare, deşi ar putea fi într-o stare între paranteze corespunzătoare).

Ceea ce s-a spus despre convingeri este valabil şi pentru intenţii. Stările intenţionale pure – de pildă, intenţia ca termenul 'apă' să se refere la apă *într-o lume noţională* – nu fixează referinţa din lumea reală. Stările intenţionale impure – de pildă intenţia ca termenul 'apă' să se refere efectiv la apă – *presupun* abilitatea de a ne referi la apa reală.

Unii filosofi au sugerat că convingerea poate fi definită în termeni de stări pe care le numesc 'stări între paranteze' *şi* referinţă, astfel:

> John crede că zăpada este albă = *John crede că* [*zăpada este albă*]

(de pildă, zăpada este albă în lumea noţională a lui John)

iar cuvintele 'zăpadă' şi 'albă' în gândirea lui John (*sau orice alte cuvinte ar fi folosite pentru a exprima o asemenea convingere*) *se referă la zăpadă şi, respectiv, la proprietatea de a fi albă.*

Chiar dacă nu reprezintă o analiză corectă şi completă a ceea ce înseamnă a fi convins că zăpada este albă, consideraţiile de mai sus subliniază ceva corect: *convingerea presupune abilitatea de a se referi.* În exact

acelaşi mod intenţia presupune abilitatea de referire! Intenţiile nu sunt evenimente mentale care *cauzează* referirea cuvintelor: intenţiile (în sensul lor obişnuit impur) au referinţă în integralitatea lor. Explicarea referinţei în termeni de intenţie impură ar fi circulară. Problema privind modul în care stările mentale *pure* de a intenţiona, de a avea convingerea etc. pot (în manieră cauzală) să constituie sau să cauzeze referinţa rămâne nerezolvată.

Originea încurcăturii

La o primă examinare, nimic nu pare mai clar decât că cuvintele şi reprezentările noastre mentale au *referinţă*. Când cred sau spun 'pisica tocmai a ieşit', mă gândesc la pisica noastră Mitty; cuvântul 'pisică' din propoziţia pe care o cred sau o spun *se referă* la o mulţime de entităţi al cărei element este şi Mitty. Încă nu am dezlegat natura acestei relaţii 'de a fi despre' a referinţei.

Distincţia dintre lumea reală şi lumea noţională (şi distincţia corelativă dintre convingeri şi convingeri între paranteze, sau intenţii şi intenţii între paranteze) explică o parte din încurcătură. Motivul pentru care este surprinzător şi tulburător să descoperim că există 'interpretări admisibile' neintenţionate pentru limbajul nostru (unde prin interpretare admisibilă înţeleg o interpretare care satisface constrângerile operaţionale şi teoretice) este, în parte, că nu ridică nici o astfel de nedeterminare în 'lumea noţională' a vorbitorului. În lumea mea noţională, pisicile şi pisicile* sunt deosebite (de fapt, în lumea mea noţională pisicile* sunt *cireşe*). 'O pisică se

află pe covor' şi 'o pisică* se află pe covor*' pot fi echivalente logic, dar ele conţin termeni cu referinţă noţională diferită; astfel, pare ciudat să existe vreo confuzie între referinţa în lumea reală a unei convingeri şi referinţa alteia.

Dar dacă numărul pisicilor se întâmplă să fie egal cu numărul cireşelor, atunci rezultă din teoremele teoriei modelelor (cum a remarcat Quine în pasajul anterior) că există o reinterpretare a întregului limbaj care păstrează valorile de adevăr ale tuturor propoziţiilor prin schimbarea între ele a extensiunilor pentru 'pisică' şi 'cireaşă'. Prin tehnica pe care am menţionat-o, asemenea reinterpretări pot fi construite aşa încât să conserve toate constrângerile teoretice şi operaţionale (iar prin tehnicile ilustrate în exemplul pisică/pisică* interpretările pot fi aplicate şi în ce priveşte 'intensiunile' sau funcţiile care determină o extensiune în orice lume posibilă şi nu numai extensiunile din lumea efectivă). Aceasta nu contravine judecăţii pe care am făcut-o în legătură cu 'lumea noţională' sau cu sistemul convingerilor subiective din următorul motiv: faptul că în sistemul convingerilor noastre sau în 'lumea noţională' nici o pisică nu este cireaşă înseamnă că în orice interpretare admisibilă pentru sistemul convingerilor (fiecare asignare de referenţi din lumea externă pentru cuvintele noastre, pentru imagini sau pentru alte reprezentări pe care le folosim în gândire) referenţii pentru 'pisică' şi pentru 'cireaşă' trebuie să aparţină la mulţimi disjuncte. Dar disjuncţia dintre aceste mulţimi este compatibilă cu faptul (remarcabil) că ceea ce este mulţimea 'pisicilor' într-o interpretare admisibilă poate fi mulţimea 'cireşelor' într-una *diferită* dar la fel de admisibilă. Din faptul că

pisicile noţionale sunt în întregime diferite de cireşele noţionale urmează numai că pisicile reale sunt cu totul diferite de cireşele reale dacă numărul interpretărilor admisibile se reduce la una. Dacă există mai mult de o interpretare admisibilă a întregului limbaj (cum ar fi dacă interpretările admisibile sunt selectate numai prin constrângeri teoretice şi operaţionale) atunci doi termeni care se referă la mulţimi disjuncte în fiecare interpretare admisibilă pot avea aceiaşi referenţi potenţiali când sunt luate în considerare toate interpretările admisibile. Din faptul că pisicile noţionale sunt atât de diferite de cireşele noţionale nu urmează că există mulţimi determinate disjuncte de pisici-în-sine şi de cireşe-în-sine.

Modul natural în care constrângerile teoretice *şi* operaţionale conduc la interpretări admisibile ale unui sistem reprezentaţional face ca întregul proces să pară atât de neobişnuit. Asemenea constrângeri pot, într-o oarecare măsură, să determine care propoziţii dintr-un limbaj sunt adevărate şi care sunt false; astfel, referinţa nu este determinată complet de condiţiile de adevăr – între acestea rămâne o nedeterminare.

Quine, aşa cum am remarcat, doreşte să elimine nedeterminarea pur şi simplu admiţând că referinţa *este* nedeterminată. Un filosof tânăr, Hartry Field[8], a sugerat recent o perspectivă diferită. Potrivit lui, referinţa este o 'relaţie fizicalistă' de pildă, un complex de relaţii cauzale între cuvinte sau reprezentările mentale şi obiecte sau

[8] Field H., 'Tarski's Theory of Truth', *The Journal of Philosophy*, vol. 69. Punctul de vedere a lui Field este discutat în lucrarea mea *Meaning and the Moral Science*, Routledge and Kegan Paul, 1978.

mulţimi de obiecte. Este sarcina ştiinţelor empirice să determine asemenea relaţii, sugerează Field.

Totuşi, apare o problemă în legătură cu această sugestie. Să presupunem că există o *definiţie* naturalistă sau fizicalistă a referinţei, aşa cum pretinde Field. Presupunem că:

(1) *x se referă la y* dacă şi numai dacă *x se află în relaţia R cu y*

este adevărată, unde *R* este o relaţie definibilă în vocabularul ştiinţelor naturii fără a folosi termeni semantici (fără a folosi 'se referă' sau alte cuvinte care ar face definiţia circulară). Dacă (1) este adevărată şi verificabilă empiric, atunci (1) este o propoziţie care este adevărată chiar dacă teoria prin care este fixată referinţa este determinată de constrângerile operaţionale *şi* teoretice. Propoziţia (1) ar fi o parte a 'echilibrului reflectiv' sau teoria 'limitei ideale' a lumii.

Dacă referinţa este determinată numai prin constrângeri operaţionale şi teoretice, atunci referinţa pentru '*x* se află în relaţia *R* cu *y*' este ea însăşi nedeterminată aşa încât, cunoaşterea că (1) este adevărată nu este de nici un ajutor. Fiecărui model admisibil pentru limbajul obiect îi va corespunde unui model pentru metalimbaj în care (1) are loc; interpretarea pentru '*x* se află în relaţia *R* cu *y*' va fixa interpretarea pentru '*x* se referă la *y*'. Dar aceasta va fi numai o relaţie *în orice model admisibil*; nu serveşte pentru a da seama de numărul tuturor modelelor admisibile.

De bună seamă că Field nu a intenţionat să tragă o asemenea consecinţă. Field afirmă că: (a) există o relaţie

unică şi determinată între cuvinte şi lucruri sau mulţimi de lucruri şi (b) această relaţie este cea folosită ca relaţie de referire în determinarea valorii de adevăr pentru (1). Dar aceasta nu este în mod necesar conţinutul a ceea ce spunem prin (1), aşa cum s-a văzut şi rămâne o problemă cum am putea *învăţa să exprimăm* ceea ce doreşte Field să spună.

Lăsând ultima problemă deoparte să considerăm că (1), înţeleasă aşa cum doreşte Field (ca descriind o relaţie unică şi determinată între cuvinte şi referenţii lor), este adevărată. În acest caz, ce o face adevărată? Având în vedere că rămân multe tipuri de corespondenţă între cuvinte şi lucruri, chiar dacă sunt satisfăcute constrângerile, care dintre ele trebuie asociată lui *R*? În nici un caz corectitudinea empirică a lui (1), deoarece avem de-a face cu constrângerile noastre teoretice şi operaţionale. Nu, aşa cum am văzut, intenţiile noastre (mai degrabă *R* intră în determinarea intenţiilor, nu invers). Rămâne varianta că teza '*R* este referinţa' trebuie să fie *metafizic inexplicabilă*, adică să fie un fapt primitiv, un adevăr metafizic.

Acest tip de adevăr metafizic, dacă ar exista, nu trebuie să fie confundat cu adevărul 'necesar metafizic' introdus recent de Saul Kripke[9].

Teoria lui Kripke, care este strâns legată de teoria prezentată mai înainte despre referinţa termenilor naturali (termeni pentru animale, vegetale, minerale) este că, fiind dat că:

---

[9] Vezi *Naming and Necessity*, Harvard U.P., 1980.

(2) Apa este H_2O

(adică, fiind dat că (2) este adevărată în lumea actuală) şi fiind dat că vorbitorul *intenţionează* ca termenul 'apă' să se refere tocmai la acele lucruri care au acelaşi comportament şi compoziţie ca diferite mostre de apă efectivă (vorbitorii pot avea o asemenea intenţie chiar dacă vorbesc despre cazuri ipotetice sau 'lumi posibile') urmează că (2) trebuie să fie adevărată în orice lume posibilă; deoarece a descrie un lichid ipotetic care nu este H_2O, dar care are unele similarităţi cu apa, înseamnă a descrie un lichid ipotetic care *se aseamănă* cu apa şi nu a descrie o lume posibilă în care apa nu este H_2O. O asemenea propoziţie, că apa este H_2O, este 'metafizic necesară' (adevărată în toate lumile posibile); dar această 'necesitate metafizică' este explicată de către chimia lumii noastre şi de faptele lumii noastre cu privire la care vorbitorul intenţionează să se refere.

Dacă există o relaţie *R* determinată fizic (definibilă în limbajul ştiinţelor naturale în finit de multe cuvinte sau nu) care este tocmai referirea indiferent cum este descrisă această relaţie), *acest* fapt nu poate fi consecinţa intenţiei noastre de a ne referi; mai degrabă, cum am arătat de mai multe ori, ea intră în determinarea intenţiei. Teza lui Kripke că 'apa este H_2O' este adevărată în toate lumile posibile, ar putea fi corectă chiar dacă referinţa din lumea efectivă este fixată numai prin constrângeri operaţionale şi teoretice; această perspectivă presupune conceptul de referinţă, nu ne spune dacă referinţa este determinată şi ce este referinţa.

Pentru mine, a considera că unele corespondenţe intrinseci sunt tocmai referinţa (nu ca rezultat al constrângerilor operaţionale şi teoretice, sau a intenţiilor noastre, dar ca fapt metafizic *ultim*) ne conduce la o teorie magică a referinţei. Referinţa însăşi, devine, dintr-o asemenea perspectivă, ceea ce Locke numea o 'formă substanţială' (o entitate care aparţine în mod intrinsec unui anumit nume). Chiar dacă cineva este dispus să analizeze asemenea fapte metafizice inexprimabile, problemele epistemologice care însoţesc o asemenea tentativă sunt insurmontabile. Astfel, prin presupunerea unei lumi a entităţilor mental şi discursiv independente (care este supoziţia perspectivei analizate) apar, aşa cum am văzut, multe 'corespondenţe' diferite care reprezintă candidaţi pentru relaţia de referire (infinit de multe, de fapt, dacă în univers există infinit de multe lucruri). Chiar cerinţa ca (1) să fie adevărată sub orice noţiune de adevăr corespunde cu alegerea metafizică a unei relaţii 'reale' a referirii care nu exclude vreunul dintre aceşti candidaţi, dacă (1) este acceptabil empiric (acceptabil în circumstanţele constrângerilor operaţionale şi teoretice), după cum s-a văzut. În acest caz, există nenumărat de multe adevăruri 'metafizic posibile' de forma '*R* este relaţia *reală* (selectată metafizic) pentru referinţă'. Dacă susţinătorul acestui punct de vedere acceptă ca o asemenea perspectivă nu este în întregime corectă şi că referinţa poate fi aleasă metafizic fără a fi *determinată* (selectarea metafizică pentru *R* poate admite o pluralitate de interpretări) atunci se poate concepe că perspectiva constrângerilor opeaţionale-plus-teoretice este metafizic corectă de fapt! Iată motivul pentru care relaţia de

referire nu poate fi un fapt metafizic: *x* se referă la *y în cel puţin un model admisibil* M. Să notăm că toate aceste teorii metafizice în număr nedefinit sunt compatibile cu *aceleaşi* propoziţii ca fiind adevărate, cu aceeaşi 'teorie asupra lumii' şi cu metodologia optimă pentru determinarea adevărului!

Două doctrine filosofice

Problemele pe care le-am discutat pun în faţa noastră două doctrine filosofice (sau două temperamente filosofice). Conţinutul şi consecinţele acestora au legătură cu subiectul nostru: problema 'creierelor în container' nu ar fi de interes decât ca un fel de paradox logic, dacă nu ar fi legată de diferenţa dintre aceste două perspective filosofice.

Prima doctrină este realismul metafizic. Din punctul ei de vedere lumea constă dintr-o totalitate fixă de obiecte independente-de-mental. Există un singur adevăr şi o singură descriere completă a 'cum este lumea'. Adevărul presupune o formă de relaţie de corespondenţă între cuvinte sau semne gândite şi lucrurile externe sau mulţimile de lucruri externe.

O voi numi doctrina *externalistă*, deoarece punctul său de vedere este cel al lui Dumnezeu.

Doctrina pe care o voi susţine nu are un nume bine stabilit. A apărut mai târziu în istoria filosofiei şi chiar şi astăzi naşte confuzii cu alte doctrine. Mă voi referi la ea spunându-i perspectiva *internalistă*, deoarece are drept caracteristică teza că întrebarea '*ce obiecte alcătuiesc lumea*?' are sens numai *în interiorul* unei teorii sau descrieri. Mulţi filosofi 'internalişti', chiar dacă nu toţi, susţin că există mai multe teorii sau descrieri 'adevărate' ale lumii. 'Adevărul', din perspectivă internalistă, este o formă (idealizată) de acceptabilitate raţională – o formă de coerenţă ideală a convingerilor noastre între ele şi cu experienţele noastre, *care la rândul lor sunt reprezentate în sistemul nostru de convingeri* – şi nu o corespondenţă cu 'stări de lucruri' independente-de-mental ori discursiv independente de discurs. Nu există nici un 'ochi al lui Dumnezeu' pe care l-am putea imagina sau cunoaşte; există numai diferite puncte de vedere ale persoanelor actuale reflectând interese şi scopuri variate, regăsite în teoriile şi descrierile pe care le acceptă. ('Teoria adevărului coerenţă, non-realism, verificaţionism, pluralism, pragmatism sunt termeni care se aplică perspectivei internaliste; dar oricare dintre aceşti termeni au conotaţii care sunt inacceptabile, dobândite prin dezvoltarea lor istorică).

Filosofii internalişti nu sunt afectaţi de ipoteza 'creierelor în container'. Pentru aceştia, 'lumea creierelor în container' este doar o *poveste*, o simplă construcţie lingvistică şi nu o lume posibilă. Ideea că o asemenea poveste ar fi adevărată într-un univers, într-un fel de realitate paralelă, presupune ochiul lui Dumnezeu încă de la început, cum se poate lesne arăta. Pentru că, *din al cărui punct de vedere poate fi spusă povestea*? De bună seamă că *nu*

din perspectiva vreunei creaturi sensibile *din* lume. Nici din perspectiva vreunui observator dintr-o altă lume care interacţionează cu lumea noastră, pentru că, prin definiţie, 'lumea' include orice interacţionează într-un fel sau altul cu lucrurile pe care le conţine. Dacă *tu*, de pildă, ai fi un observator care nu are creierul în container, şi ai putea observa alte creiere aflate în container atunci nu am avea de-a face cu o lume în care *toate* fiinţele sensibile ar fi creiere în container. Astfel, presupunerea că ar putea exista o lume în care *toate* fiinţele sensibile sunt creiere în container poate fi făcută numai din afara acestei lumi, din perspectiva asupra adevărului a ochiului lui Dumnezeu, sau, mai exact, dintr-o perspectivă în care adevărul ar fi absolut independent de orice observator din acea lume.

Pentru filosoful externalist, pe de altă parte, ipoteza că suntem creiere în container nu poate fi eliminată prea lesne. Deoarece adevărul unei teorii nu depinde de corespondenţa cu lumea aşa cum este observată de către observatorii din ea (adevărul nu este *relaţional*), ci de corespondenţa cu lumea în sine. Problema care apare pentru filosoful externalist este că orice relaţie de corespondenţă de care depind adevărul şi referinţa nu poate fi valabilă logic pentru el, dacă este creier în container. Astfel, dacă *suntem* creiere în container, nu putem *gândi* că suntem astfel, cu excepţia punerii între paranteze [noi suntem creiere în container]; iar acest gând între paranteze nu are condiţii de referinţă care l-ar face *adevărat*. Prin urmare, nu este posibil să fim creiere în container.

Să admitem 'teoria magică a referinţei'. De pildă, am putea presupune că unele legături oculte – să le numim 'legături noetice'[1] – leagă cuvintele sau gândurile-semne de

[1] Termenul 'legături noetice' mi-a fost sugerat de Zemach.

referinţa lor. În acest caz, nu există probleme. Creierul din container poate crede *cuvintele* 'Eu sunt un creier în container' iar cuvântului 'container' să-i corespundă (cu ajutorul legăturilor noetice) containerele reale din exterior, iar cuvântului 'în' să-i corespundă (cu ajutorul legăturilor noetice) relaţia spaţială de conţinere. Dar o asemenea perspectivă nu poate fi susţinută. Nici unul dintre filosofii contemporani nu ar fi de acord cu ea deoarece realistul modern admite teoria adevărului corespondenţă fără a crede în legături noetice (sau fără a crede în obiecte care se identifică singure[2] – obiecte care corespund intrinsec unui cuvânt sau gând-semn mai degrabă decât altuia) aşa încât creierele în container ar fi o ciudăţenie pentru el.

Cum am văzut, problema este următoarea: există aceste obiecte în exterior. Aici este mintea/creierul cu funcţia gândire/calcul. Cum anume, simbolurile gânditorului (sau cele ale minţii/creierului) intră într-o corespondenţă unică cu obiectele din exterior?

Răspunsul acceptat de externaliştii de astăzi este că, în vreme ce nu este nevoie ca semnele să corespundă în mod *necesar* unei mulţimi de lucruri mai degrabă decât alteia, conexiunile *contextuale* între semne şi lucrurile externe (în particular, conexiunile cauzale) ar fi în măsură să explice natura referinţei. Dar acest mecanism nu funcţionează. De pildă, convingerile mele privind electronii se datorează *cărţilor*. Dar rostirile mele ale cuvântului 'electron' deşi au, într-un anume înţeles, legătură cu textele cărţilor, nu se *referă* la asemenea texte. Obiectele care sunt cauza dominantă a convingerilor mele conţin semnul dar nu referinţa acestui semn.

---

[2] Termenul 'obiecte care se identifică singure' este din *Substance and Sameness* de D. Wiggins (Blackwell, 1980).

Externalistul va răspunde că un cuvânt ca 'electron' nu este conectat cu textul cărţii printr-un lanţ cauzal de forma adecvată. (Dar cum putem determina care lanţ cauzal are forma corectă decât dacă am fi deja capabili să recunoaştem referinţa?)

Pentru un internalist, cum sunt şi eu, situaţia este diferită. Din perspectivă internalistă, semnele nu corespund intrinsec obiectelor independent de modul în care sunt utilizate şi de către cine. Un semn care este utilizat într-un mod particular de către o anume comunitate de utilizatori poate corespunde obiectelor particulare *în interiorul schemei conceptuale a acestor utilizatori.* 'Obiectele' nu există independent de schemele conceptuale. *Noi* decupăm lumea în obiecte când introducem o schemă sau alta de descriere. Deoarece atât obiectele cât *şi* semnele sunt *interne* relativ la o schemă de descriere, putem explica de ce se potrivesc.

Într-adevăr, este trivial să spunem că orice cuvânt se referă *în interiorul* limbajului căruia îi aparţine prin folosirea cuvântului însuşi. La ce se referă 'iepure'? De bună seamă, la iepuri! La ce se referă 'extraterestru'? La extratereştri (dacă există vreunul).

De bună seamă, externalistul este de acord că extensiunea lui 'iepure' este mulţimea iepurilor şi extensiunea lui 'extraterestru' este mulţimea extratereştrilor. Dar el nu consideră asemenea judecăţi ca spunând ce *este* referinţa. Pentru el, a determina ce *este* referinţa, respectiv care este natura corespondenţei între cuvinte şi lucruri, reprezintă problema importantă. (*Cât* de importantă, am văzut în capitolul precedent). Pentru mine, sunt puţine de pus despre ce este referinţa în interiorul sistemului

conceptual, altceva decât tautologii. Ideea că legătura cauzală este necesară este respinsă de faptul că 'extraterestru' se referă la extratereştri indiferent dacă am avea vreo legătură cauzală cu extratereştrii sau nu!

Filosoful externalist va replica, totuşi, că ne putem referi la extratereştri chiar dacă nu am interacţionat niciodată cu ei (după câte ştim) deoarece am interacţionat cu *tereştri* şi avem situaţii empirice ale relaţiei 'nu provine de pe aceeaşi planetă' şi situaţii pentru 'fiinţă inteligentă'. Acum putem defini un extraterestru ca o fiinţă inteligentă care provine de pe o altă planetă decât aceea a tereştrilor. În plus, 'nu provine de pe aceeaşi planetă' poate fi analizat în termenii 'nu provine din acelaşi loc' şi 'planetă' (care pot fi analizaţi mai departe). În acest fel, externaliştii elaborează condiţiile pentru a avea loc o conexiune 'reală' (bunăoară, cauzală) cu orice situaţie la care suntem în măsură să ne referim şi condiţii în care termenii de *bază* se referă la tipuri de lucruri (şi relaţii) prin mijlocirea conexiunilor reale. Folosind termenii de bază în combinaţii complexe putem, spun ei, să construim expresii descriptive care se referă la tipuri de lucruri fără a avea conexiune reală cu ele, ba chiar fără ca acestea să existe (cum sunt extratereştrii).

De fapt, chiar cuvinte simple, cum sunt 'cal' şi 'iepure' includ în extensiunea lor obiecte faţă de care nu avem vreo legătură cauzală (de pildă, caii şi iepurii viitori ori caii şi iepurii care nu au interacţionat niciodată cu oamenii). Când folosim cuvântul 'cal' ne referim nu numai la caii cu care am avut o legătură reală, ci la toate lucrurile *de acelaşi fel*.

În acest loc trebuie să observăm că 'de acelaşi fel' nu are sens în afara sistemului categorial care spune ce proprietăţi contează şi ce proprietăţi nu contează ca similarităţi. Într-un mod sau altul, orice poate fi 'de acelaşi fel' cu orice. Întreaga poveste despre modul în care ne referim la unele lucruri în virtutea faptului că sunt legate printr-un lanţ cauzal cu noi şi cu alte lucruri în virtutea faptului că sunt 'de acelaşi fel' şi cu celelalte 'prin descriere' este atât falsă, cât şi inutilă. Asemănarea dintre caii cu care nu am interacţionat niciodată şi caii cu care am interacţionat este dată de faptul că primii sunt, la urma urmelor, *cai*. Formularea problemei de către realistul metafizic presupune că există toate obiectele în ele însele şi noi folosim un fel de lasou pentru a prinde unele dintre ele (cele faţă de care avem legături reale) şi aplicăm cuvântul ('cal') pentru a le acoperi nu numai pe cele prinse cu lasoul ci şi pe celelalte, care nu au fost prinse pentru că ar fi prea îndepărtate în spaţiu şi timp sau din alte motive. 'Soluţia' acestei pseudo-probleme – 'realist' metafizică – este a considera cuvintele ca acoperind *automat* nu numai obiectele prinse cu lasoul, ci şi obiectele *de acelaşi fel* – de acelaşi fel *în sine*. În acest caz, lumea conţine, la urma urmelor, obiecte care se autoidentifică deoarece tocmai asta înseamnă că *lumea* şi nu cugetătorii clasifică obiectele în tipuri.

Lumea constă din 'obiecte care se autoidentifică' dar nu în sensul valabil pentru externalist. Dacă, aşa cum aminteam, 'obiectele' sunt mai degrabă produse decât descoperite, aşa cum producerea invenţiilor conceptuale este un factor 'obiectiv' al experienţei, factorul independent faţă de voinţa noastră, atunci, de bună seamă, obiectele aparţin în mod intrinsec unor categorii; deoarece aceste categorii sunt determinate de uneltele folosite pentru

construirea versiunii lumii în care obiectele sunt pe primul loc. Dar acest tip de 'obiecte care se autoidentifică' nu este independent-de-mental; dimpotrivă, externalistul doreşte să creadă că lumea constă din obiecte care *simultan* sunt independente-de-mental şi se autoidentifică. O asemenea variantă nu poate avea loc.

Internalism şi relativism

Internalismul nu este un relativism facil care afirmă 'merge orice'. A nega că întrebarea privind 'potrivirea' conceptelor noastre cu lucruri necontaminate de conceptualizare ar avea sens, este una; dar a susţine că orice sistem conceptual este la fel de bun cu oricare altul, este cu totul altceva. Dacă cineva chiar crede aceasta şi este suficient de nebun pentru a avea un sistem conceptual care îi spune că sărind pe fereastră ar putea zbura, ar constata, dacă ar supravieţui, slăbiciunea punctului de vedere de mai sus. Internalismul nu neagă că există *intrări* empirice ale cunoaşterii; cunoaşterea nu este un basm care să nu aibă alte constrângeri decât coerenţa *internă*; dar neagă că ar exista intrări *care nu ar fi ele însele modelate de conceptele noastre*, prin vocabularul pe care îl folosim pentru a le descrie sau orice intrare *care admite numai o singură descriere independent de orice alegere conceptuală*. Chiar descrierile propriilor senzaţii, atât de utilizate ca punct de plecare pentru generaţii întregi de epistemologi (care servesc ca senzaţii în sine), sunt afectate de alegerile conceptuale ale celui care le posedă. Intrările pe care se bazează cunoaşterea sunt contaminate conceptual, dar intrările contaminate sunt mai bune decât nici o intrare.

Dacă tot la ce ne raportăm sunt intrările contaminate tot este mai bine decât nimic.

O judecată sau un întreg sistem de judecăţi – o teorie sau o schemă conceptuală – este acceptabilă raţional, în mare parte, datorită coerenţei sale, este vorba de coerenţa convingerilor 'teoretice' între ele şi faţă de convingerile cu preponderenţă empirice şi, de asemenea, coerenţa între convingerile empirice şi cele teoretice. Concepţiile noastre despre coerenţă şi acceptabilitate sunt, din perspectiva pe care o prezint, strâns legate de psihologia noastră. Ele depind de biologia şi cultura noastră; ele nu sunt 'independente de valori'. Dar ele *sunt* concepţiile noastre şi sunt concepţii despre ceva real. Ele definesc un tip de obiectivitate, *obiectivitatea pentru noi* chiar dacă nu este obiectivitatea metafizică a ochiului lui Dumnezeu. Obiectivitatea şi raţionalitatea din perspectivă umană sunt tot ce avem; ele sunt mai bune decât nimic.

Respingerea ideii că există o perspectivă 'externalistă' coerentă, o teorie care este adevărată în sine, separat de toate posibilităţile de observare, nu se *identifică* cu acceptabilitatea raţională. Adevărul nu poate fi confundat cu acceptabilitatea raţională datorită unui motiv important; adevărul este o proprietate a judecăţilor care nu poate fi pierdută, în vreme ce acceptabilitatea raţională poate fi pierdută. Judecata 'Pământul este plat' era acceptabilă raţional acum 3 000 de ani, cu aproximaţie; dar nu este raţional acceptabilă astăzi. Ar fi eronat să se afirme că 'Pământul este plat' era adevărată acum 3 000 ani; deoarece asta ar însemna că Pământul şi-a schimbat forma. De fapt, acceptabilitatea raţională este relativă atât faţă de timp, cât şi faţă de persoană. În plus, acceptabilitatea raţională este

graduală; adevărul este uneori utilizat gradual (bunăoară, uneori spunem *'Pământul este o sferă' este aproximativ adevărat*); dar 'gradual' aici înseamnă *acurateţea* judecăţii şi nu gradul de acceptare sau justificare.

Aceasta nu arată, în opinia mea, că externalistul are dreptate în ultimă instanţă, ci că adevărul este o *idealizare* a acceptabilităţii raţionale. Vorbim ca şi când dacă există condiţii epistemologice ideale, putem numi o judecată 'adevărată' dacă ar fi justificată sub asemenea condiţii. 'Condiţiile epistemice ideale' sunt, de bună seamă, aidoma 'absenţei frecării': nu putem atinge în mod real aceste condiţii ba chiar este sigur că ne aflăm destul de departe de ele. Dar, deşi absenţa frecării nu poate fi atinsă, are sens să vorbim despre asemenea situaţii deoarece le putem aproxima în suficientă măsură.

Poate s-ar părea că explicarea adevărului în termeni de justificare sub condiţii ideale înseamnă punerea unei noţiuni clare în termenii uneia vagi. Dar 'adevărat' nu este atât de clar cât ar părea din exemple precum 'zăpada este albă'. În orice caz, nu încerc să dau o definiţie formală a adevărului, ci o elucidare informală a acestuia.

Analogia cu absenţa frecării ne conduce la cele două idei cheie ale idealizării teoriei adevărului: (1) adevărul este independent de justificarea aici şi acum, dar nu este independent de orice justificare. A susţine că o judecată este adevărată înseamnă a susţine că poate fi justificată; (2) se aşteaptă ca adevărul să fie stabil sau 'convergent'; dacă atât o judecată cât şi negaţia sa ar putea fi 'justificate' chiar când condiţiile ar fi ideale, nu există nici un motiv să gândim o asemenea judecată ca *având* o valoare de adevăr.

Teoria 'asemănării'

Teoria că adevărul este corespondenţă pare cu siguranţă cea mai firească. Înainte de Kant era practic imposibil de găsit un filosof care să fie împotriva teoriei adevărului corespondenţă.

Michael Dummett a trasat recent[3] distincţia între *non-realişti* (cam ceea ce eu numesc 'internalişti') şi *reducţionişti*, pentru a sublinia că reducţionistii pot fi realişti metafizici şi subscriu teoriei adevărului corespondenţă. Reducţionismul consideră că există o clasă de aserţiuni (de pildă, aserţiuni despre evenimente mentale) care sunt 'făcute adevărate' prin fapte din afara clasei. De exemplu, fapte privind comportamentul 'fac adevărate' aserţiuni despre evenimente mentale, conform cu un anume tip de reducţionism. Un alt exemplu de perspectivă *reducţionistă* este cea a lui Berkeley, că numai minţile şi senzaţiile lor sunt 'reale', aşa încât propoziţiile despre mese şi scaune cât şi despre alte 'obiecte materiale' sunt adevărate despre senzaţii.

Dacă o perspectivă este reducţionistă cu privire la aserţiunile de un anumit fel, dar numai pentru a susţine teoria adevărului corespondenţă pentru propoziţiile din clasa *redusă*, atunci perspectiva este la bază realist metafizică. Un adevărat non-realist este non-realist în toate privinţele.

Confuzia între filosofii reducţionişti şi cei non-realişti este făcută adesea, dar Dummett pune lucrurile la punct; dezacordul dintre cele două tipuri de filosofi priveşte *ceea ce este real* şi nu concepţia asupra adevărului. Dacă evităm o asemenea eroare, atunci afirmaţia pe care

[3] Tezele lui Dummett se găsesc în 'What is a Theory of Meaning I, II', *Truth and Other Enigmas* (Harvard, 1980).

am făcut-o, că este imposibil să găsim, înainte de Kant, filosofi care nu sunt realişti metafizici, cel puţin în ce priveşte aserţiunile de *bază* nereductibile, pare mult mai plauzibilă.

Cea mai veche teorie a adevărului corespondenţă datează de aproximativ 2 000 ani fiind atribuită de filosofii antici şi medievali lui Aristotel. Nu sunt sigur că Aristotel a susţinut-o; dar a sugerat-o prin expresiile pe care le-a folosit. O voi numi *teoria referinţei-similaritate,* deoarece susţine că relaţia între reprezentările minţilor noastre şi obiectele externe la care se referă este de *similaritate.*

Această teorie, aidoma teoriilor moderne, foloseşte ideea de reprezentare mentală. Imaginea în minte a lucrurilor externe a fost numită *fantomă* de către Aristotel. Relaţia dintre fantomă şi obiectul extern, în virtutea căreia fantoma reprezintă obiectul extern în minte, este (conform lui Aristotel) că fantoma *împărtăşeşte aceeaşi formă* cu obiectul extern. Deoarece fantoma şi obiectul extern sunt similare (împărtăşesc aceeaşi formă), mintea, când are fantome corecte, surprinde forma corectă a obiectelor externe[4].

Aristotel însuşi spune că fantomele *nu* împărtăşesc cu obiectele alte proprietăţi, precum *roşeaţa* (roşeaţa din mintea noastră nu este aceeaşi proprietate cu roşeaţa obiectului) care poate fi percepută printr-un simţ, dar împărtăşesc proprietăţi precum *lungimea* ori *aspectul* care sunt percepute prin mai mult decât un singur simţ (care sunt 'sensibil comune' în opoziţie cu 'sensibil singulare').

În secolul al XVII-lea teoria similarităţii a început să fie restricţionată mai mult decât o făcuse Aristotel. Astfel,

[4] *De Anima*, Cartea III, Cap. 7 şi 8.

Locke şi Descartes au susţinut că, în cazul calităţilor 'secundare', cum ar fi culorile sau textura, ar fi absurd să se presupună că proprietatea imaginii mentale este *literal* aceeaşi proprietate cu aceea a lucrului fizic. Locke era un corpuscular, respectiv, un apărător al teoriei atomice a materiei şi asemenea fizicienilor moderni credea că roşeaţa prezentă în reprezentarea mentală a unei bucăţi de stofă nu este o proprietate simplă a stofei, ci un complex de proprietăţi dispoziţionale sau o 'putere', respectiv puterea de a provoca senzaţii de un anume tip (senzaţii care se manifestă ca 'roşu subiectiv' în limbajul psihofizicienilor). Această putere, la rândul său, îşi găseşte o explicaţie care nu era cunoscută pe vremea lui Locke, în microstructura bucăţii de stofă care absoarbe sau reflectă în mod selectiv lumina de anumite lungimi de undă. (O asemenea explicaţie era anticipată de Newton). Dacă spunem că 'roşu' înseamnă a avea o anumită microstructură în cazul unei bucăţi de stofă, atunci caracterul subiectiv al roşului poate fi dat prin evenimentul din mintea mea sau chiar din creierul meu care are loc când am o senzaţie de roşu subiectiv şi *nu* implică prezenţa în mintea mea a ceva roşu. Proprietăţile lucrurilor fizice şi proprietăţile mentalului care produc roşul sunt diferite. O bucată roşie de stofă şi o imagine de roşu nu sunt similare literal. Ele nu împărtăşesc aceeaşi formă.

Pentru asemenea proprietăţi (aspect, mişcare, poziţie) care sunt privite de filosofia corpusculară ca fundamentale şi ireductibile, Locke tinde să păstreze teoria referinţei-similaritate. (De fapt, unii dintre comentatorii de astăzi a lui Locke pun în discuţie acest fapt; dar Locke spune că există o 'similaritate' între idee şi obiect în cazul

calităţilor primare şi că nu există nici o similaritate între ideea de *roşu* sau de *căldură* şi roşeaţa sau căldura din obiect[5]. Interpretarea lui Locke pe care am menţionat-o era unanim recunoscută de contemporanii săi şi în secolul al XVIII-lea).

Turul de forţă a lui Berkeley

Berkeley descoperă o consecinţă neplăcută a teoriei referinţei similaritate: aceasta implică faptul că nu există nimic în afara entităţilor mentale ('spirite şi ideile lor', sau minţi şi senzaţiile lor). De regulă nu se observă că premisa pe care se bazează Berkeley – teoria similarităţii – nu era pur şi simplu învăţată sau preluată de la Locke, ci avea în vedere teoria anterioară şi, încă, cea de o sută de ani după aceea; deja am remarcat cât de veche era această teorie.

Raţionamentul lui Berkeley este foarte simplu. Argumentul obişnuit împotriva similarităţii în cazul calităţilor secundare (argumentul din relativitatea percepţiei) merge la fel de bine şi în cazul proprietăţilor primare. Lungimea, aspectul sau mişcarea unui obiect sunt percepute diferit de observatori diferiţi şi de acelaşi observator în împrejurări diferite. A întreba dacă o *masă* are aceeaşi lungime cu imaginea *mea* a mesei sau aceeaşi lungime cu imaginea *ta* este o întrebare absurdă. Dacă masa are 90 cm lungime şi eu o disting în întregul ei, atunci imaginea are 90 cm lungime? O asemenea întrebare nu are sens. Imaginile mentale nu au lungime *fizică*. Ele nu pot fi comparate cu ajutorul unităţii standard de la Paris. Lungimea fizică şi lungimea subiectivă

[5] *An Essay Concerning Human Understanding*, Cartea II, Cap. VIII.

trebuie să fie tot atât de diferite între ele cum sunt, una față de alta, roșeața fizică și cea subiectivă.

Concluzia lui Berkeley poate fi formulată astfel: *nimic nu poate fi similar unei senzații sau imagini decât o altă senzație sau imagine*. Având în vedere acest fapt, cât și presupunerea (care nu era pusă la îndoială) că mecanismul referinței este similaritatea între 'ideile' noastre (sau imagini sau 'fantome') și ceea ce ele reprezintă, urmează că nici o idee (imagine mentală) nu poate reprezenta și nu se poate referi la altceva decât o altă imagine sau senzație. Pot fi gândite, concepute, ne putem referi la, numai obiecte fenomenale. Iar dacă nu poți gândi ceva înseamnă că nu poți gândi că există. Numai dacă tratăm discursul despre obiecte materiale ca derivat în întregime din discursul despre regularitățile din senzațiile noastre, acesta devine neinteligibil.

Tendința, în timpul său și mai târziu, de a-l vedea pe Berkeley ca pe un pervers nesănătos, ca pe un ins aproape scandalos, chiar dacă strălucitor, era determinată de neacceptabilitatea concluziei sale că materia nu există în realitate (cu excepția constructelor din senzații) și nu cu privire la vreuna dintre premise. Dar faptul că cineva a putut deriva o asemenea concluzie inacceptabilă din teoria similarității a generat o criză în filosofie. Filosofii care nu doreau să-l urmeze pe Berkeley în idealismul său subiectiv au ajuns la abordări diferite ale referinței.

Viziunea lui Kant asupra cunoașterii și adevărului

Doresc să spun că filosofia lui Kant este interpretată cel mai bine prin prisma a ceea ce am numit 'internalism' sau 'realism internalist' în ce privește concepția despre

adevăr, deşi el nu a recunoscut aproape niciodată că ar fi urmărit un asemenea scop.

Mai întâi, Kant a privit idealismul subiectiv al lui Berkeley ca inacceptabil (mai mult decât a spus-o el însuşi) şi, de asemenea, realismul cauzal – punctul de vedere că noi percepem în mod direct numai senzaţii şi *inferăm* obiectele materiale printr-un fel de inferenţă problematică, i se pare la fel de inacceptabil. O perspectivă prin care, faptul că în faţa mea există o masă atunci când scriu aceste pagini este numai o ipoteză dubioasă, este 'scandaloasă' pentru Kant.

În al doilea rând, Kant a văzut clar că argumentul lui Berkeley funcţionează: el a observat că acest argument depinde de teoria referinţei-similaritate şi că respingerea argumentului lui Berkeley cere respingerea acestei teorii. Aici am atribuit lui Kant o perspectivă pe care acesta nu o redă explicit (într-adevăr, a vorbi despre 'referinţă' ca relaţie între semnele mentale şi lucrurile pentru care ele stau este o manieră foarte recentă, deşi problema relaţiei dintre semnele mentale şi lucruri este foarte veche). Vom vedea că tezele lui Kant au exact efectul amintit asupra teoriei referinţei similaritate.

Sugerez o cale de a-l interpreta pe Kant care poate fi de ajutor, deşi este numai o primă aproximare pentru o interpretare corectă. Kant a acceptat punctul de vedere al lui Berkeley că argumentul din relativitatea percepţiei se aplică şi aşa-numitelor calităţi primare ca şi celor secundare, dar a dat un răspuns diferit celui al lui Berkeley. Răspunsul lui Berkeley, după cum s-a văzut, era de a şterge distincţia între calităţile primare şi cele secundare şi de a considera senzaţiile ca entităţile de bază la care ne referim. Locke considera calităţile secundare ca puteri sau

proprietăţi – *de natură nespecificată* – care permit obiectului să ne afecteze într-un anumit fel. A spune despre ceva că este roşu sau că este cald, sau furios, înseamnă a spune că se află într-o anume *relaţie cu noi* şi nu că aşa se arată ochiului lui Dumnezeu.

Eu sugerez că (la o primă aproximaţie) modul de interpretare a lui Kant este de a extinde ceea ce spune Locke despre calităţile secundare asupra *tuturor* calităţilor – cele simple, cele primare sau secundare (într-adevăr există o oarecare distincţie între ele)[6].

Dacă *toate proprietăţile sunt secundare*, ce urmează? Urmează că *tot* ce putem spune despre un obiect are forma: este astfel încât ne afectează în cutare sau cutare mod. Nu am putea spune *nimic mai mult* despre vreun obiect descris 'în sine' independent de efectul asupra *noastră*, ca fiinţe raţionale şi având constituţia noastră biologică. Urmează, de asemenea, că nu putem presupune vreo similaritate între ideea noastră

---

[6] Kant face un rezumat al doctrinei sale în *Prolegomena*: 'Încă multă vreme înaintea epocii lui Locke, dar mai ales după aceea, a fost unanim recunoscut şi acceptat că, în pofida existenţei reale a lucrurilor externe, se poate spune despre o mulţime dintre predicatele lor că nu aparţin acestor lucruri în sine, ci doar fenomenelor lor şi că ele nu au o existenţă proprie în afara reprezentării noastre. Printre aceste predicate se numără temperatura, culoarea, gustul etc. Nu văd însă nici un argument serios care ar putea fi adus împotriva considerării celorlalte calităţi ale corpurilor – care sunt numite *primare*, şi anume întinderea, locul şi, în genere, spaţiul cu tot ce ţine de el (impenetrabilitatea sau materialitatea, forma etc.) – în rândul simplelor fenomene. Şi după cum nu poate fi numit idealist acela care consideră culorile nu drept proprietăţi ale obiectului în sine, ci drept modificări ale simţului văzului, tot aşa nu ar putea fi numită idealistă doctrina mea doar pentru că socotesc că încă şi mai multe proprietăţi, ba chiar *toate acele proprietăţi care constituie intuiţia unui corp* aparţin doar fenomenului său...' (Kant I.: *Prolegomene*, Ed. Ştiinţifică şi Enciclopedică, Bucureşti, 1987, p. 85–86 – *n. trad.*)

despre un obiect şi orice realitate independentă-de-mental care poate fi, în ultimă instanţă, responsabilă pentru experienţa pe care o avem despre acel obiect.

Kant nu se îndoieşte că există o realitate independentă-de-mental ; pentru el acesta este un postulat al raţiunii. El se referă la elementele acestei realităţi independente-de-mental cu diferiţi termeni: lucru în sine (*Ding an sich*); obiect noumenal sau *noumen*; colectiv, *lumea noumenală*. Dar nu putem exprima nici o concepţie despre lucrurile noumenale; chiar noţiunea de lume noumenală este un fel de limită a gândirii (*Grenz-Begriff*) mai degrabă decât un concept clar. Astăzi noţiunea de lume noumenală este percepută ca fiind un element metafizic nenecesar în sistemul lui Kant. (Dar poate Kant are dreptate: poate nu putem stabili cu ajutorul gândirii că există vreo bază independentă-de-mental a experienţei noastre chiar dacă încercarea de a vorbi despre aceasta conduce la un non-sens).

În acelaşi timp, a vorbi de obiecte 'empirice' obişnuite nu înseamnă a vorbi de lucrurile în sine, ci numai de lucrurile pentru noi.

Adevărata subtilitate este că I. Kant priveşte toate aceste perspective ca aplicându-se senzaţiilor ('obiecte ale simţului intern') *la fel de bine* ca obiectelor externe. Aceasta poate părea straniu: cum corespunde o idee unei senzaţii? Dar Kant ajunge la ceva mai profund.

Să presupunem că am o senzaţie *E*. Să presupunem că eu *descriu E*, bunăoară prin aserţiunea '*E* este senzaţia de *roşu*'. Dacă 'roşu' înseamnă *ca acesta* atunci întreaga aserţiune înseamnă '*E* este *ca acesta*' (cu privire la *E*), bunăoară, *E este ca E* – şi nu este făcută nici o judecată. Cum ar spune Wittgenstein, descrierea se reduce, de fapt, la un mormăit. Pe de altă parte, dacă *roşu* este un *clasificator*,

dacă afirm că *E aparţine aceleiaşi clase* cu senzaţiile pe care le-am numit 'roşu' altcândva, atunci judecata merge dincolo de datul imediat, dincolo de aparenţă şi presupune o referire implicită la alte senzaţii pe care nu le am la momentul de faţă şi la *timp* (care, conform lui Kant nu este ceva noumenal, ci mai degrabă o formă în care aranjăm 'lucrurile pentru noi')[7]. Dacă senzaţiile pe care le am la momente diferite şi pe care le clasific drept *senzaţii de roşu* sunt toate în mod real (noumenal) similare este o problemă care nu are sens; dacă ele apar ca fiind similare (de pildă dacă îmi *amintesc* senzaţiile anterioare ca similare cu aceasta şi *anticipez* că senzaţiile viitoare vor fi astfel clasificate încât să pară similare cu modul în care aceasta va fi reamintită) atunci ele sunt similare-pentru-mine.

Kant repetă de mai multe ori şi în termeni diferiţi că obiectele simţurilor interne nu sunt real transcendentale (noumenale) ci ele sunt ideal transcendentale (lucruri pentru noi) şi că ele nu sunt mai mult sau mai puţin cognoscibile direct decât aşa numitele obiecte 'externe'. Senzaţiile pe care le numesc 'roşu' nu pot fi comparate direct cu obiectele noumenale pentru a vedea dacă acestea au aceeaşi proprietate noumenală mai mult decât ar putea fi comparate direct obiectele pe care le numesc 'bucăţi de aur' cu obiectele noumenale pentru a vedea dacă au aceeaşi proprietate noumenală.

Motivul pentru care 'toate proprietăţile sunt secundare' este, într-o primă aproximaţie pentru Kant, următorul: 'Toate proprietăţile sunt secundare' (sau *toate*

---

[7] Aici am procedat deliberat anacronic explicând pe Kant printr-un exemplu luat din *Philosophical Investigation*. Dar exemplul lui Wittgenstein îşi are rădăcinile în Kant: Hegel, scriind la scurt timp după Kant şi conştient de doctrina kantiană, arată că orice judecată, chiar despre impresiile simţurilor, trebuie să treacă dincolo de ce este 'datul' pentru a *fi* o judecată.

proprietăţile sunt puteri) sugerează că, a spune că un scaun este făcut din pin, sau din altceva, înseamnă a atribui o putere (dispoziţia de a apărea ca făcut din pin pentru noi) unui obiect noumenal; a spune că scaunul este maro, înseamnă a atribui o altă putere aceluiaşi obiect noumenal etc. Un asemenea punct de vedere ar face să corespundă câte un obiect noumenal fiecărei reprezentări kantiene, fiecărui lucru pentru noi. Dar Kant *neagă* explicit o asemenea situaţie. Acesta este momentul la care se desparte de teoria adevărului corespondenţă.

Kant însuşi nu spune că renunţă la teoria adevărului corespondenţă. Dimpotrivă, el afirmă că adevărul este 'corespondenţa dintre judecată şi obiectul său'. Dar aici este ceea ce Kant numeşte 'definiţia nominală a adevărului'. În viziunea mea, identificarea acesteia cu ceea ce înţelege realismul metafizic prin adevăr corespondenţă ar fi o eroare gravă. Pentru a spune despre Kant dacă susţine teoria adevărului corespondenţă în varianta realismului metafizic trebuie să vedem care este concepţia sa despre ceea ce numeşte 'obiect' al unei judecăţi empirice.

Din punctul de vedere a lui Kant, orice judecată despre obiecte externe sau interne (lucruri fizice sau entităţi mentale) spune că lumea noumenală ca întreg este astfel încât judecata respectivă este descrierea pe care ar construi-o o fiinţă raţională (cineva care are natura noastră raţională) care are informaţii valabile pentru o fiinţă cu organe de simţ (o fiinţă cu o natură sensibilă). În *acest* sens, judecata atribuie o putere. Dar puterea este atribuită *lumii noumenale în întregul său*; nu putem gândi că deoarece există scaune şi cai şi senzaţii în reprezentările noastre, acestora le corespund în lumea noumenală scaune şi cai noumenali, cât şi senzaţii noumenale. *Nu există o corespondenţă biunivocă între lucrurile pentru noi şi lucrurile în sine.* Kant, nu numai că respinge orice similaritate între ideile noastre şi lucrurile în sine; el

respinge şi noţiunea de izomorfism abstract. Aceasta înseamnă că nu există nici o teorie a adevărului corespondenţă în filosofia sa.

Ce este atunci o judecată adevărată? Kant crede că avem o cunoaştere obiectivă: noi cunoaştem legile matematicii, legile geometriei, legile fizicii şi multe judecăţi despre obiecte individuale – obiecte empirice, lucruri pentru noi. Folosirea termenului 'cunoaştere' şi folosirea termenului 'obiectiv' conduce la aserţiunea că *există o noţiune de adevăr*. Dar ce este adevărul dacă nu este corespondenţă cu lucrurile în sine?

După cum am văzut, singurul răspuns care poate fi extras din ceea ce spune Kant este următorul: un element de cunoaştere (o 'judecată adevărată') este o judecată pe care o fiinţă raţională ar accepta-o dacă ar avea o experienţă suficientă de tipul pe care o asemenea fiinţă ar fi în măsură să îl aibă. 'Adevărul' în orice alt sens ne este inaccesibil şi nu poate fi conceput de către noi. *Adevărul este, în ultimă instanţă, ce ni se potriveşte.*

Alternativa empiristă

Până acum, argumentul nostru lasă încă posibilitatea să se evite renunţarea la teoria adevărului corespondenţă şi teoria referinţei-similaritate prin *restrângerea la senzaţii şi imagini*. Mulţi filosofi au continuat să creadă, chiar după Kant, că similaritatea *este* mecanismul prin care suntem în măsură să avem idei care se referă la propriile noastre *senzaţii* (şi, deşi este o opinie controversată, la ale altora ca noi) şi că acesta este punctul de pornire al referirii din perspectivă epistemologică.

Pentru a arăta că un asemenea mecanism nu funcţionează, să ne reamintim că nucleul argumentului lui

Berkeley era constatarea că nimic nu se aseamănă cu o 'idee' (senzaţie sau imagine) cu excepţia altei 'idei', bunăoară nu poate exista asemănare între fizic şi mental. Ideile noastre se aseamănă numai cu entităţi mentale, dar ele nu pot fi asemenea 'materiei' conform lui Berkeley.

Aici intervine o importantă supoziţie falsă. De fapt, *orice este similar cu orice chiar într-o infinitate de aspecte.* De exemplu, senzaţia mea când scriu la maşină în acest moment şi ceasul din buzunarul meu sunt similare sub raportul unor proprietăţi (senzaţia care apare chiar acum şi prezenţa ceasului în buzunarul meu chiar acum) sunt *efecte ale acţiunilor mele trecute*; dacă nu aş fi stat jos să scriu nu aş fi avut asemenea senzaţii; iar ceasul nu ar fi în buzunarul meu dacă nu l-aş fi pus acolo. Atât senzaţia, cât şi ceasul există în secolul al XXI-lea. Atât senzaţia cât şi ceasul sunt descrise în română etc. Numărul similarităţilor care pot fi găsite între *oricare* două obiecte este limitat numai de imaginaţie şi timp.

Într-un context particular, 'similaritatea' poate avea o semnificaţie mai restrictivă, de bună seamă. Dar tocmai întrebarea '*A* şi *B* sunt similare?', când nu este precizat tipul de similaritate pe care îl avem în vedere, este o întrebare la care nu se poate răspunde.

Din acest fapt simplu am tras deja consecinţa că teza potrivit căreia similaritatea este mecanismul referinţei conduce la regresie la infinit. Să presupunem, pentru a prelua un exemplu de la Wittgenstein, că cineva încearcă să inventeze un 'limbaj privat', un limbaj care se referă la propriile sale senzaţii. El îşi concentrează atenţia asupra senzaţiei *X* şi introduce semnul *E* pe care intenţionează să-l aplice exact acelor entităţi care sunt identice calitativ cu *X*. De fapt, el intenţionează ca *E* să fie aplicat tuturor şi numai acelor entităţi care sunt *similare* cu *X*.

Dacă intenţia sa este numai aceasta – dacă el nu specifică *reperul* faţă de care ceva să fie similar cu *X* pentru a cădea sub clasificarea *E* – atunci intenţia sa este irealizabilă, deoarece, aşa cum am văzut, orice este similar cu *X* într-o privinţă sau alta.

Dacă, pe de altă parte, el precizează reperul; dacă are gândul că *o senzaţie este E dacă şi numai dacă este similară cu X în privinţa R* atunci, când are acest gând, el este în măsură să se refere la senzaţiile pentru care încearcă să introducă termenul *E* şi la proprietăţile relevante ale acestor senzaţii! Dar cum ajunge să facă acest lucru? (Dacă răspundem 'prin concentrarea atenţiei asupra a două senzaţii *Z*, *W*, şi prin gândul că două senzaţii sunt similare cu privire la *R* dacă şi numai dacă sunt similare ca şi *Z* cu *W*' atunci se ajunge la regres la infinit.)

Dificultatea în ceea ce priveşte teoria referinţei-similaritate este aceeaşi cu dificultatea privind 'lanţul cauzal de tip adecvat' despre care am vorbit. Dacă tocmai am spus 'Cuvântul "cal" se referă la obiecte care au proprietatea a cărei apariţie îmi provoacă ocazia de a rosti "în faţa mea se află un cal"', atunci dificultatea care apare este că există prea multe asemenea proprietăţi. De pildă, fie *H–A* (apariţia unui cal) proprietatea pentru toate situaţiile perceptuale care cer răspunsul 'în faţa mea se află un cal' (chiar când avem de-a face cu o iluzie), dar 'cal' nu se referă la situaţiile cu această proprietate, ci mai degrabă la anumite animale. Prezenţa unui animal cu proprietatea de a aparţine unei categorii naturale şi prezenţa unei situaţii conceptuale cu proprietatea *H–A* sunt ambele legate de rostirea 'în faţa mea se află un cal' prin lanţuri cauzale. De fapt, apariţia cailor în Epoca de piatră este legată de rostirea mea 'în faţa mea se află un cal' printr-un lanţ cauzal. Aşa cum există *prea* multe similarităţi pentru ca referinţa să fie

o simplă problemă de similaritate, există prea multe lanţuri cauzale pentru ca referinţa să fie reductibilă la acestea.

Pe de altă parte, dacă spun 'cuvântul "cal" se referă la obiecte care au o proprietate care este legată de rostirea "în faţa mea se află un cal" în anumite ocazii printr-un lanţ cauzal adecvat' atunci apare problema că, dacă sunt în măsură să arăt care este lanţul cauzal adecvat trebuie *deja* să fiu în stare să mă refer la tipurile de obiecte şi proprietăţi care sunt legate prin acest lanţ cauzal. Dar ce anume ne permite să atingem această performanţă?

Concluzia nu este că nu există termeni care pot fi descrişi prin teoria similarităţii mai mult decât că nu există termeni care pot fi descrişi prin apelul la lanţuri cauzale de un anumit tip. Concluzia este că nici teoria similarităţii nici cea a conexiunii cauzale nu pot fi mecanismul referirii.

Wittgenstein despre 'urmarea unei reguli'

Să ne oprim la exemplul pe care l-am menţionat în trecere, al celui care încearcă să stabilească reperul *R* (reperul faţă de care senzaţiile trebuie să fie similare cu *X* dacă ele sunt clasificate corect ca *E*) prin a spune sau a gândi că două lucruri sunt similare relativ la *R* numai în cazul în care ele sunt similare în felul în care *Z* şi *W* sunt similare. Această încercare dă greş deoarece oricare două lucruri *Z* şi *W* sunt similare în moduri diferite (de fapt, în infinit de multe moduri). Să încercăm să precizăm relaţia de similaritate dând un număr finit de exemple aşa cum încercăm să determinăm o funcţie de numere naturale

având în vedere 1 000 (sau 1 000 000) de valori: există întotdeauna infinit de multe funcţii care sunt conforme cu o mulţime finită de valori, dar care diferă prin valorile care nu sunt pe listă.

Această observaţie este legată de o altă opinie a lui Wittgenstein din *Investigaţii filosofice* care a fost amintită la sfârşitul capitolului I. Orice semne de care ne putem da seama prin introspecţie, sau 'reprezentări' pe care suntem capabili să le invocăm în legătură cu un concept, nu pot determina sau constitui *conţinutul* acelui concept. Wittgenstein face această constatare în faimoasa secţiune privind 'urmarea unei reguli', să spunem regula 'adaugă unu'. Chiar dacă două specii din două lumi posibile (eu formulez argumentul într-o terminologie care nu aparţine lui Wittgesntein!) au aceleaşi semne mentale în legătură cu regula 'adaugă unu', este posibil ca modul în care o pun în practică să fie diferit; şi practica este cea care fixează interpretarea: semnele nu se interpretează prin ele însele, după cum am văzut. Chiar dacă cineva aparţinând celeilalte specii înţelege relaţia '*A* este succesorul lui *B*' (*A* = *B*+1) în acelaşi fel cu noi în ce priveşte o mulţime numeroasă dar finită de cazuri (2 este succesorul lui 1, 3 este succesorul lui 2, ..., 999 978 este succesorul lui 999 977) încă am putea fi în dezacord privind interpretarea lui 'succesor' pentru unele cazuri viitoare. (Chiar dacă celălalt ar fi de acord cu noi la nivel teoretic privind ceea ce se înţelege prin 'succesor', ar putea avea o interpretare diferită a întregii teorii, aşa cum arată teorema Löwenheim-Skolem).

Acest fapt are o relevanţă imediată pentru filosofia matematicii, ca şi pentru filosofia limbajului. Mai întâi de toate este o problemă de *finitism*: practica umană, actuală sau

potenţială, este finită. Chiar dacă spunem că putem, nu vom ajunge niciodată să numărăm perpetuu. Dacă pot fi *extensiuni divergente ale practicii noastre, atunci sunt posibile interpretări divergente chiar ale secvenţei numerelor naturale* – practica noastră, sau reprezentările noastre mentale etc., nu impun un 'model standard' unic pentru şirul numerelor naturale. Noi suntem tentaţi să credem în unicitatea modelului deoarece trecem de la 'am putea continua să numărăm' la 'o maşină ideală ar putea continua să numere' (sau 'o *minte* ideală ar putea continua să numere'); dar a vorbi despre maşini ideale (sau minţi) este altceva decât a vorbi despre maşini sau persoane efective. A vorbi despre ce ar putea face o maşină ideală ar însemna să vorbim în interiorul matematicii, nu să fixăm o interpretare a matematicii.

În acelaşi mod, Wittgenstein susţine că a vorbi despre 'similaritate' şi 'aceeaşi senzaţie' sau 'aceeaşi experienţă' înseamnă a vorbi în interiorul unei teorii psihologice; nu putem fixa interpretarea acelei teorii psihologice. *Aceasta*, interpretarea teoriei şi terminologiei psihologice, este fixată prin practica noastră actuală de standardele actuale de corectitudine şi incorectitudine.

În *Modalităţi de a construi lumea*[8] Nelson Goodman face o observaţie apropiată: o noţiune despre ce sunt faptele perceptuale 'în realitate' independent de modul în care sunt conceptualizate sau de descrierile pe care le dăm şi care ni se par corecte nu are rost. Astfel, după discutarea tezei psihologului Kolers că un număr disproporţionat de ingineri şi fizicieni nu sunt în stare să-şi dea seama de mişcarea aparentă produsă de lumini care se aprind succesiv în diferite poziţii, Goodman comentează (p. 92):

[8] Publicat de Hackett, 1978.

Mai mult, faptul că un subiect susţine că vede două lumini care se aprind distinct la distanţe şi intervale atât de scurte încât cei mai mulţi observatori văd o mişcare a unui singur punct luminos, este asemănător cu a spune că privind un scaun vedem o îngrămădire de molecule sau că tăblia unei mese o vedem rotundă atunci când privim oblic. Dacă un observator este adept al realităţii dincolo de aparenţe, poate considera aparenţa mişcării ca un semn că există două lumini care se aprind succesiv sau să considere aparenţa ovalului tăbliei mesei ca un semn că aceasta este rotundă; în ambele cazuri semnele pot fi sau deveni atât de transparente încât să vedem obiectele şi evenimentele fizice dincolo de ele. Când observatorul determină vizual ce se află în faţa sa, numai cu greu se poate spune că percepţia sa este eronată. Atunci, care este sursa neînţelegerilor asupra a ce spune o persoană că vede? Cum anume pot fi reformulate instrucţiunile de observare pentru a preveni neînţelegerile? Solicitaţi cuiva să nu apeleze la experienţa anterioară şi să evite orice conceptualizare şi îl veţi lăsa fără grai; atunci când vorbim trebuie să folosim cuvinte.

Surprinderea 'formelor' şi asocierea empirică

Un platonist sau un neo-platonist ar fi rezolvat problema într-un mod mult mai simplu: ar fi spus că, atunci când avem o senzaţie particulară percepem şi o formă universală, că mintea are capacitatea de a surprinde proprietăţile în sine şi nu numai concretizări ale lor. Un asemenea filosof ar considera ca nominalistă perspectiva lui Wittgenstein sau Goodman, deoarece aceştia refuză posibilitatea surprinderii directe a vreunei forme, ceea ce îi duce la dificultăţi în ce priveşte teoria similarităţii.

Cu toate acestea, misterioasa putere de a 'surprinde forme' este cu greu acceptabilă ca o soluţie potrivită, deşi ar părea că şi din perspectiva de aici se încearcă ceva asemănător. Proprietăţile lucrurilor intră în *explicaţii* cauzale; când avem o senzaţie care cere răspunsul 'aceasta este o senzaţie de roşu', răspunsul meu este cauzat în parte de faptul că senzaţia are o *proprietate*. Unii filosofi sunt atât de nominalişti încât tind să respingă existenţa unor entităţi precum 'proprietăţile'; dar ştiinţa însăşi nu ezită să vorbească liber despre proprietăţi. Am putea să nu spunem, atunci când privatul lui Wittgenstein (cel care doreşte să inventeze un limbaj privat) are în vedere *X* şi spune '*E*', că faptul care a cauzat răspunsul *E* a fost o interacţiune cauzală în care este prezentă o *proprietate* şi că acea proprietate (oricare ar fi fost) priveşte 'similaritatea' relevantă pe care alte senzaţii trebuie să le aibă cu *X* pentru a fi clasificate corect ca *E*?

Observaţia că a vorbi despre proprietăţi este legitim ştiinţific este corectă, dar nu ajută la reabilitarea platonismului. Noi interacţionăm cu proprietăţile numai prin intermediul concretizărilor acestora, iar aceste concretizări sunt întotdeauna ale mai multor proprietăţi simultan. Nu există nici o interacţiune cu proprietăţile 'în sine'. Apelul la proprietăţi asociate cauzal cu o senzaţie nu ne poate conduce la noţiunea de formă (unică) a senzaţiei aşa cum face filosofia platonică.

Lucrurile stau astfel: când am senzaţia de albastru, am senzaţia de *albastru* şi, în plus, am senzaţia proprietăţii complexe de existenţă după cum sunt clasificate de mine sub diferite etichete verbale. Mai simplu, dobândirea acestei senzaţii nu constituie 'surprinderea' *uneia* dintre aceste proprietăţi. Dacă vrem să evidenţiem proprietatea asociată cu senzaţia sau cu eticheta verbală numai într-un singur mod,

ajungem la vechea noastră cunoştinţă, lanţul cauzal de un tip adecvat.

Pentru a stabili acest fapt, să observăm mai întâi de toate că, atunci când experienţa perceptuală totală conduce la răspunsul 'Am senzaţia de albastru', nu am întotdeauna *dreptate*. Eu însumi am avut experienţa referirii la 'omul cu un pulover albastru' de două sau trei ori înainte să mi se atragă atenţia că este un pulover verde. Nu vreau să spun că puloverul *părea* albastru; mi-am dat seama că am descris greşit puloverul în momentul în care cealaltă persoană m-a atenţionat. (Nu am deseori ocazia să spun 'Am senzaţia de albastru' dar, într-un asemenea caz, aş fi repetat probabil de două sau trei ori până când cineva – mirat că am senzaţia de albastru când indic spre ceva care este evident verde – mi-ar fi atras atenţia şi mi-aş fi dat seama de eroare). Acestea arată că proprietatea care determină răspunsul 'Am senzaţia de albastru', sau un altul, *nu este aceeaşi proprietate* cu proprietatea de a fi o senzaţie de albastru, sau o senzaţie de orice alt tip.

De obicei, asemenea situaţii sunt descrise printr-o expresie de felul 'de a sta pe limbă'. Mi se pare o terminologie nepotrivită. Cuvântul 'verde' putea să-mi fi stat pe limbă şi să mă surprind, fără să vreau, spunând 'albastru'. *Aici* ar fi vorba de a-mi fi stat pe limbă. Dar în cazul descris nu mi-am dat seama de eroare până când cineva mi-a pus la îndoială raportul* (altfel s-ar putea să nu fi realizat niciodată greşeala).

O altă explicaţie ar fi că, atunci când am spus 'albastru', înţelegeam verde. Dar ar trebui să fie limpede că,

---

* (enunţul observaţional sau 'de protocol' – *n. trad.*)

atunci când spunem ceva, înţelesul nu rămâne în minte. A spune 'eu înţeleg' verde înseamnă tocmai a spune că instantaneu accept corecţia (şi mă amuz când îmi dau seama ce am spus). Aceasta numai repetă ce s-a întâmplat, nu este o explicaţie.

Oricare ar fi explicaţia (poate o scăpare în unitatea de procesare a vorbirii din creierul meu), întocmai cum proprietatea *A–H* despre care am amintit determină raportul: 'În faţa mea se află un cal' chiar într-o ocazie când în jurul meu nu se găseşte nici un cal, la fel există un complex de proprietăţi al ansamblului minţii mele care determină enunţul 'Am senzaţia de albastru', când nu am senzaţia de albastru (sau, oricum, aş nega că o am dacă aş fi atenţionat). Nici un mecanism de asociere empirică nu este perfect. Dacă decidem să spunem despre cineva că are o senzaţie de albastru ori de câte ori o senzaţie determină, pentru acesta, raportul 'Am o senzaţie de albastru' (sau care determină raportul şi un asemenea raport nu pare eronat la o a doua examinare) atunci pentru teoria psihologică populară, ba chiar psihologia ştiinţifică, acestea ar fi un criteriu pentru a stabili că cineva are senzaţia de albastru, chiar dacă, în multe împrejurări, senzaţia nu ar fi albastru. Mai mult, aşa cum arată Wittgenstein, pornind de la un asemenea criteriu, *ceea ce mie mi se pare corect este corect* – în acest caz, distincţia între a face un raport corect privind o senzaţie şi a face un raport care mi se pare corect ar trebui abandonată. Poate ar trebui să o abandonăm sau cel puţin să o estompăm; poate, cum pare să sugereze Goodman, problema dacă avem în realitate senzaţia pe care credem că o avem, nu are sens, cu excepţia unor cazuri speciale, cum ar fi situaţia în care cineva ne atrage atenţia; dar abandonarea acestei distincţii nu este posibilă pentru realistul metafizic deoarece realismul metafizic s-a

constituit tocmai privind distincţia între ceea ce este real şi ce crede cineva că se întâmplă.

Se poate greşi în ceea ce priveşte calitatea senzaţiilor trecute?

Altă cale de a aborda problema noastră este prin punerea întrebării: 'S-ar putea greşi întotdeauna în ce priveşte calitatea senzaţiilor trecute?' Potrivit teoriei similarităţii, răspunsul este 'da'. Conform acestei teorii senzaţiile mele anterioare fie sunt, fie nu sunt similare cu senzaţiile pe care le descriu *acum* prin etichete verbale precum 'senzaţie de roşu', 'durere' etc. şi dacă ele sunt sau nu sunt similare este o cu totul altă problemă decât dacă eu le-am clasificat corect atunci când le-am avut, sub aceleaşi etichete verbale. Poate că lumea este astfel încât ceea ce am numit 'senzaţie de roşu' în minutele pare numărate de la începutul erei noastre, pentru minutele impare este identic cu ceea ce numesc 'senzaţie de verde', dar memoria le confundă. Senzaţia clasificată sub eticheta verbală 'senzaţie de roşu' cu un minut în urmă nu ar mai fi similară cu senzaţia pe care o clasific astfel acum.

Există ceva care nu se potriveşte cu acest scenariu. Există un motiv pentru care sensul în care 'nu am observat niciodată o anume senzaţie' este prea puternic: dacă eu consider senzaţia mea de roşu legată la momente diferite de diverse semnale sau de apariţii fizice (foc, semnalul de stop etc.) atunci am succes în acţiunile mele. Clasele de similaritate 'greşite' (clasa care adună laolaltă toate senzaţiile pe care le numesc senzaţii de roşu în ciuda faptului că ele nu au în realitate aceeaşi 'calitate') reprezintă mijlocul prin care aş avea un succes mai mare în rezolvarea problemelor puse de acţiunile mele. Dar atunci, asemenea clase sunt într-adevăr *greşite*?

Dacă nu presupun că noţiunea de similaritate este autointerpretantă, atunci acest caz poate fi descris ca unul în care relaţia numită 'similaritate' de un observator extern, diferă de relaţia pe care noi o numim 'similaritate'. Din această perspectivă, ipoteza că ne înşelăm 'în realitate' cu privire la senzaţiile trecute, cade: din punct de vedere *internalist* nu există vreo noţiune inteligibilă de senzaţii care la momente diferite sunt 'similare' dincolo de standardele de acceptabilitate raţională.

Din nou despre teoria adevărului corespondenţă

Până acum cititorul trebuie să fie deja convins că teoria referinţei-similaritate trebuie abandonată. Dar de ce acest fapt ar trebui să atragă renunţarea la teoria adevărului corespondenţă? Chiar dacă noţiunea de 'similaritate' între conceptele noastre şi între referinţa lor nu funcţionează, nu ar putea exista un fel de izomorfism abstract, sau, chiar dacă nu un izomorfism în înţelesul deplin al cuvântului, vreun fel de corespondenţă abstractă între concepte şi lucrurile (independente-de-mental) care există în lume? Oare adevărul nu ar putea fi definit pornind de la o relaţie de izomorfism sau de corespondenţă?

Necazul cu o astfel de sugestie nu este că între concepte şi lucruri nu ar putea exista vreo corespondenţă, ci există *prea* multe asemenea corespondenţe. Pentru a selecta numai o singură corespondenţă între cuvinte sau semne mentale şi lucrurile independente-de-mental ar trebui să avem acces referenţial la lucrurile independente-de-mental. Nu am

putea să ne oprim la o singură corespondenţă între două lucruri decât în mod forţat; nu am putea determina corespondenţa dintre conceptele noastre şi obiectele noumenale fără a avea acces la acestea din urmă.

Un argument în acest sens este următorul. Uneori teoriile incompatibile pot fi intertraductibile. De exemplu, dacă fizica lui Newton ar fi adevărată, atunci orice fizică ar putea fi descrisă în două moduri: în termeni de particule care acţionează la distanţă în interiorul spaţiului vid (cum descrie Newton acţiunea gravitaţiei) sau în termeni de particule care acţionează prin intermediul câmpurilor. De exemplu, ecuaţiile lui Maxwell, care descriu comportamentul câmpului electromagnetic, sunt echivalente matematic cu o teorie în care apare numai acţiunea la distanţă între particule, respectiv atracţia invers proporţională cu pătratul distanţei, care se propagă nu instantaneu, ci mai degrabă cu viteza luminii ('potenţiale întârziate'). Teoria câmpului a lui Maxwell şi teoria potenţialului întârziat sunt incompatibile metafizic deoarece fie există, fie nu există agenţi cauzali (cum sunt câmpurile) care mediază interacţiunile particulelor separate (cum ar spune un realist). Dar cele două teorii sunt intertraductibile matematic. Astfel, în cazul în care corespondenţa cu lucrurile noumenale care ne conduce la adevăr există, am putea defini o altă corespondenţă care ar face cealaltă teorie adevărată. Dacă ceea ce face o teorie adevărată este o corespondenţă abstractă (nu contează care) atunci unele teorii incompatibile ar fi adevărate împreună.

Pentru un internalist, acest fapt nu este de neacceptat: de ce să nu existe teorii la fel de coerente chiar dacă incompatibile conceptual? Dacă adevărul nu este o corespondenţă (unică) atunci se deschide posibilitatea pluralismului. Dar realistul metafizic urmăreşte să salveze

noţiunea de ochi al lui Dumnezeu printr-o teorie a adevărului unic.

Nu numai că poate exista corespondenţă între obiecte şi (ceea ce considerăm a fi) teorii incompatibile (aceleaşi obiecte pot fi ceea ce logicienii numesc 'model' pentru teorii incompatibile) dar chiar dacă teoria şi obiectele care există sunt determinate (dacă sunt în număr infinit) rămân infinit de multe moduri *diferite* în care acelaşi obiect poate fi folosit pentru a obţine un model pentru o teorie dată. Aşa se exprimă în limbaj matematic faptul intuitiv că selectarea unei corespondenţe între două domenii necesită o cale de acces independentă spre ambele domenii.

Am examinat o teorie care a dominat două mii de ani. Faptul că a persistat atât timp în variate forme în ciuda contradicţiilor interne şi a obscurităţilor pe care le conţine arată puterea de atracţie a ochiului lui Dumnezeu. Kant, primul care a susţinut că această atracţie nu este suficientă, considera că este încastrată în natura noastră raţională (el sugerea sublimarea acestei atracţii 'totalizatoare' prin încercarea de a realiza 'cel mai mare bine în lume', prin reconcilierea ordinii morale şi ordinii empirice într-un sistem perfect de instituţii sociale şi relaţii individuale). Prezenţa continuă a acestui impuls natural dar insuficient este, probabil, cauza adâncă pentru monismele şi dualismele false care au proliferat în cultura noastră; dar acestea pot fi ocolite prin înlăturarea ochiului lui Dumnezeu.

Minte şi corp

Paralelism, interacţionism, identitate

În secolul al XVII-lea marii filosofi Descartes, Spinoza şi Leibniz au înţeles importanţa relaţiei dintre minte şi corp. Într-o primă formă, această relaţie a fost abordată încă de Platon; dar ea devine cu atât mai mult o problemă odată cu apariţia fizicii moderne. În secolul al XVII-lea, oamenii începeau să fie conştienţi că lumea fizică este *închisă cauzal*. Închiderea cauzală a fost exprimată cel mai bine în termenii fizicii newtoniene: nici un corp nu se mişcă decât ca rezultat al acţiunii unei *forţe*. Forţele pot fi descrise complet prin numere: sunt suficiente trei numere pentru a determina direcţia şi încă un număr

pentru a determina mărimea forţei. Acceleraţia produsă de o forţă are exact direcţia forţei, iar mărimea acceleraţiei poate fi calculată din masa corpului şi mărimea forţei potrivit primei legi a lui Newton $F = ma$. Când asupra unui corp acţionează mai mult decât o forţă, rezultanta poate fi determinată prin regula paralelogramului.

Este important de subliniat cât de diferită apare o asemenea fizică, apelând la numere şi algoritmi precişi de calcul, faţă de gândirea calitativă a Evului Mediu. În gândirea medievală aproape orice exercita o influenţă asupra a orice altceva. (Cuvântul 'influenza' a supravieţuit modului de gândire medieval. Spiritele rele erau concepute ca experte în influenţare – *questa influenza* în italiană – încât puteau provoca boli pe calea aerului.) Potrivit acestui mod de a vedea lucrurile, nu este surprinzător că mintea poate influenţa corpul.

Pe vremea filosofilor amintiţi, metoda matematică începuse să dea la o parte acest mod de gândire. Noua perspectivă nu a fost dezvoltată în întregime până la Newton, dar o contribuţie însemnată o are Descartes cu regula paralelogramului şi, mai înainte, Leonardo da Vinci. Aceşti gânditori au văzut că fizica poate fi dezvoltată într-un mod asemănător cu cel în care este prezentată astăzi. Ei şi-au dat seama că fizica se ocupă cu forţe şi mişcare, şi respinge stilul calitativ de explicare. Mai mult, ei au înţeles că lumea mecanică are o logică proprie, are un 'program', cum am spune noi şi urmează acest program până ce intervine o perturbare.

Acestor gânditori li se părea că evenimentele mentale ar putea însemna unul sau două lucruri. (1) Ele ar putea fi *paralele* evenimentelor fizice, bunăoară, evenimente în creier. Modelul este o pereche de ceasuri sincronizate: corpul este un ceas care poate fi afectat şi care se îndreaptă

fără speranţă spre moarte, aşa cum întreaga lume fizică parcurge drumul de la Creaţie la Judecata de Apoi (sau la colapsul gravitaţional într-o versiune mai nouă). Evenimentele mentale au drumul lor propriu şi cumva, probabil prin intervenţia providenţei divine, se petrec astfel încât evenimentul cerebral *B* apare întotdeauna ca răspuns la senzaţia *S*. (2) Acestea ar putea *interacţiona* cu evenimentele fizice. Evenimentele mentale ar putea cauza evenimente cerebrale şi viceversa.

Teza lui Descartes, vestit prin perspectiva interacţionistă, care sugerează că mintea poate influenţa materia când materia este foarte eterică (şi că, în fapt, acţionează asupra materiei prin glanda pineală), era cel puţin o speculaţie sau, mai mult, o continuare a doctrinelor medievale[1]. În modul de gândire anterior, mintea era înfăţişată ca acţionând prin intermediul unui 'spirit' asupra materiei, în vreme ce spiritul nu era gândit în *totalitate* imaterial. 'Spiritul' era tocmai un tip de entitate intermediară pe care filosofii medievali aveau tendinţa să o introducă între orice termeni vecini dintr-o serie de tipuri naturale. El era ca un gaz capabil de o apăsare uşoară. Câtă vreme 'spiritul' este lăsat deoparte şi mintea este considerată imaterială, presiunea minţii asupra materiei, chiar eterice, în glanda pineală, apare de neînţeles. Nimeni nu ar putea vizualiza un asemenea proces.

Cea mai naivă formă a interacţionismului concepe mintea ca un fel de stafie capabilă să locuiască în diferite corpuri (dar fără a modifica în acest fel ce gândeşte, simte, îşi reaminteşte şi cum se manifestă personalitatea, judecând după spusele cărţilor despre reîncarnare şi 'reamintirea

[1] *The Discarded Image*, de C.S. Lewis, în special Cap. VII, sec. F, pentru o descriere a perspectivei medievale (Cambridge, 1964).

vieţilor anterioare), ba chiar fiind capabilă să existe fără corp (continuând să gândească, să simtă, să-şi reamintească şi să-şi manifeste personalitatea). Această versiune, care aduce puţin a superstiţie, este vulnerabilă obiecţiei că există o evidenţă copleşitoare (o parte era deja cunoscută în secolul al XVII-lea) că funcţii precum gândirea, simţirea sau memoria au loc în *creier*. În ipoteza amintită, nu este clar de ce creierul ar trebui să fie atât de complicat. Pentru funcţia de simplă 'timonă', creierul uman este prea complex.

Pentru a evita obiecţiile aduse de ştiinţă, interacţionişti mai evoluaţi, cum este Descartes, susţin că mintea şi creierul formează o *unitate esenţială*. Într-un anume mod, unitatea minte-creier este cea care gândeşte, simte, îşi reaminteşte şi îşi manifestă personalitatea. Asta înseamnă că termenul 'minte' nu semnifică mintea, ci unitatea minte-creier. Această doctrină, care ne spune că există ceva alcătuit din două substanţe atât de diferite precum mintea şi creierul aflate într-o unitate esenţială, rămâne extrem de obscură.

Alternativa este la fel de stranie. Ce determină un eveniment mental să însoţească un eveniment cerebral? Un filosof al secolului al XVII-lea sugera că evenimentele mentale ar putea fi în mod actual identice cu cele cerebrale sau cu alte evenimente fizice, iar acesta a fost Spinoza. Sugestia, în forma dată de autor, este că un eveniment, precum durerea pe care o resimt într-o ocazie anume ar putea fi identic cu faptul că creierul meu este într-o stare *B* tocmai în aceeaşi ocazie. (Voi exprima această perspectivă spunând că proprietăţile ca eu să simt o anumită durere şi creierul meu să se afle într-o stare *B* sunt identice. Prefer să vorbesc în această manieră deoarece cred că avem o teorie mai elaborată a proprietăţilor decât a evenimentelor, dar

ideea poate fi exprimată în ambele moduri. Teza, formulată în această terminologie, este că proprietatea unei persoane de a avea senzaţia Q ar putea fi *aceeaşi proprietate* cu a creierului de a fi în starea *B*.) În această formă, ideea era sugerată de Diderot, de pildă, în secolul al XVIII-lea şi a devenit populară prin anii 1940 şi 1950. Atunci, materialismul şi teoria identităţii încep să fie luate în serios pentru prima dată, fiind adoptată teza lui Spinoza (sau Spinoza mai puţin aspectul teologic şi metafizic): ne confruntăm cu o singură lume reală şi faptul că încă nu ştim cum anume stările de a avea durere, a auzi sunete, a avea senzaţii vizuale sunt stări ale creierului nu înseamnă că aceasta nu ar putea fi.

Prima formă contemporană a teoriei identităţii a fost elaborată în mai multe scrieri şi a fost mai bine cunoscută prin filosoful australian J. J. C. Smart. Cea mai importantă teză este că o senzaţie, să spunem de albastru, este identică cu o anumită stare neurofiziologică. O variantă a acesteia a fost sugerată mai întâi de către mine şi am numit-o *funcţionalism*[2]. Pentru funcţionalist există într-adevăr o identitate, dar nu proprietăţile creierului pe care le lua în considerare Smart sunt celălalt termen al identităţii. Conform funcţionalismului, creierul are proprietăţi care *nu* sunt de natură fizică.

Acum, ce înţeleg prin a spune că creierul are proprietăţi care nu sunt fizice? Înţeleg proprietăţi care nu sunt *definibile în termenii fizicii sau chimiei*. Dacă pare straniu ca un sistem care este fizic să aibă proprietăţi care nu sunt fizice, să ne oprim la un computer. Un computer are,

---

[2] Lucrarea lui N. Block: *Readings in Philosophy of Psychology* (Harvard 1980) conţine o colecţie excelentă de articole despre funcţionalism. Studiul meu a fost retipărit în *Mind, Language and Reality, Philosophical Papers*, vol. 2, (Cambridge, 1975).

desigur, multe proprietăţi fizice: are o anumită greutate, are un număr de circuite etc. Pe de altă parte, are proprietăţi economice, de pildă, are un anumit preţ; în plus are proprietăţi funcţionale, cum ar fi, un anumit program. Acest ultim tip de proprietăţi nu este fizic, *în sensul că poate fi realizat de un sistem independent de compoziţia sa metafizică sau ontologică.* Un spirit fără corp ar putea executa un anumit program, un creier ar putea funcţiona după un anume program, o maşină urmează un program, iar organizarea funcţională a tuturor acestora ar fi aceeaşi deşi materia şi elementele din care sunt alcătuite sunt diferite.

Proprietăţile psihologice manifestă aceeaşi caracteristică; aceeaşi proprietate psihologică (furia) poate fi o proprietate a mii de specii diferite care pot avea o fizică şi o chimie cu totul diferite (unele din aceste specii pot fi extraterestre, ba chiar roboţii vor fi într-o zi furioşi). Funcţionalismul este mai plauzibil decât monismul în secolul al XX-lea, iar cea mai plauzibilă teorie care nu consideră mintea şi materia ca două substanţe separate sau două domenii aparte de proprietăţi este cea care identifică proprietăţile psihologice cu cele funcţionale.

Astăzi mai sunt înclinat să susţin că această teorie este corectă; sau cel puţin este corectă în descrierea *naturalistă* a relaţiei minte/corp. Există alte descrieri, 'mentaliste', ale acestei relaţii care şi ele sunt corecte, dar nu sunt reductibile la ceea ce numim 'Natură' (bunăoară, noţiunile de 'raţionalitate', 'adevăr', 'referinţă' *aparţin* versiunii mentaliste). Voi spune câte ceva despre ele mai târziu (Capitolul 6). Acest fapt nu mă descurajază: deoarece, aşa cum a arătat Nelson Goodman, una dintre trăsăturile atractive ale non-realismului este că lasă posibilitatea mai multor alternative corecte asupra lumii. Totuşi, rămân atras de ideea unei singure versiuni corecte, care este cea naturalistă, în care formele gândirii, imaginile,

senzaţiile etc. sunt caracteristici funcţionale ale unor fenomene fizice; aş dori să abordez dificultatea cu care se confruntă funcţionalismul în ultimii ani privind caracterul calitativ al senzaţiilor. Când cineva are în vedere stări psihologice pure şi abstracte, de pildă cele rezultate prin punerea între paranteze, cum ar fi un gând considerat numai în conţinutul său noţional, sau stări difuze emoţionale, ca a fi gelos sau a fi furios, atunci identificarea acestora cu stări funcţionale ale întregului sistem pare foarte plauzibilă; dar când avem în vedere o calitate prezentă, cum ar fi experienţa unei nuanţe de albastru, identificarea nu mai este plauzibilă.

În prelegerile mele am folosit de multe ori o variantă a faimosului exemplu a 'spectrului inversat' (care apare în lucrările lui Locke[3]), unde un individ vede lucrurile astfel încât roşu îi apare albastru şi albastru îi apare roşu (respectiv, culorile sale subiective corespund culorilor negative şi nu culorilor pozitive). Prima reacţie a cuiva pus în faţa unui astfel de exemplu ar fi 'Bietul băiat, oamenilor ar trebui să le fie milă de el'. Dar cum ar cunoaşte altcineva situaţia sa? Când el vede ceva albastru, i se pare roşu, dar el numeşte acea culoare 'albastru' încă de când era copil aşa încât dacă cineva îl întreabă ce culoare are lucrul, va răspunde 'albastru'. Astfel nimeni nu şi-ar da seama.

Modificarea mea este următoarea: să ne imaginăm că spectrul se inversează la un moment dat şi *chiar ne amintim cum era înainte.* În acest caz nu există o problemă epistemologică a verificării. Ne-am trezi dimineaţa şi cerul ar părea roşu, iar un pulover roşu ar fi albastru şi toate feţele ar avea o culoare neplăcută, negativul celei obişnuite. O Dumnezeule! Acum, individul ar putea modifica modul

---

[3] *An Essay Concerning Human Understanding*, Cartea II, Cap. 32 (sec. 14).

în care se exprimă şi ar numi lucrurile care ne apar roşii 'albastre' aşa încât răspunsul la ce culoare are puloverul să fie unul normal. Numai noaptea, subiectul nostru ar ofta uneori: 'Oh, aş dori să văd culorile aşa cum le vedeam când eram copil. Culorile nu arată ca de obicei'.

În acest caz ni se pare că am putea şti ce s-a întâmplat. Unele 'fire' trebuie să se fi 'încurcat' în creier. Intrările pentru lumina albastră care foloseau o anumită cale, acum folosesc alta, iar intrările pentru lumina roşie, prima cale. Cu alte cuvinte, mecanismele cerebrale s-au schimbat între ele, afectând *stările fizice.* O stare care juca iniţial un anumit *rol funcţional* semnalizând prezenţa albastrului 'obiectiv' în mediu acum semnalează prezenţa roşului 'obiectiv' în acelaşi mediu.

Să presupunem că adoptăm următoarea teorie funcţionalistă asupra culorilor subiective: 'o senzaţie este o senzaţie de albastru (are *calitatea* descrisă *acum* în acest mod) în cazul în care senzaţia (sau evenimentul fizic corespunzător din creier) are rolul de a semnala prezenţa albastrului obiectiv în mediu'. Această teorie dă sens expresiei 'senzaţia de albastru', dar nu îi dă sensul calitativ pe care îl urmărim. Dacă rolul funcţional ar fi *identic* cu caracterul calitativ atunci nu s-ar putea constata schimbarea calităţii senzaţiei. (Pentru clarificare să ne imaginăm că după inversarea spectrului şi după ce am învăţat să o compensăm lingvistic, avem un atac amnezic care înlătură amintirile privind culorile de odinioară. În acest caz ni s-ar părea că senzaţia care acum este numită 'senzaţie de albastru' ar avea exact acelaşi rol funcţional cu senzaţia care a fost numită iniţial astfel, când, de fapt, ea are o cu totul altă calitate.) Calitatea nu pare a fi o stare *funcţională* în *acest* caz.

Cred că strategia potrivită pentru un funcţionalist în faţa unor asemenea situaţii ar fi să spună: 'Da, dar *calitatea* constă tocmai în aspectul fizic'. Ar spune că, pentru acest tip special de proprietăţi psihologice, pentru *calităţi*, vechea formă a teoriei identităţii era cea corectă. Dacă cititorul are înclinaţii materialiste, probabil ar gândi că proprietatea de a avea senzaţii este una cerebrală. Cititorii care nu au înclinaţii materialiste vor considera că proprietatea de a avea senzaţii este *corelată* cu o stare cerebrală. Probabil că cei mai mulţi oameni ar adopta una dintre aceste două teze: stările-senzaţii sunt corelate cu stări cerebrale sau stările-senzaţii sunt identice cu stările cerebrale. Cum se întâmplă adesea, problema se învârte mereu în jurul aceleiaşi întrebări: 'fiind dat că *B* este întotdeauna corelat cu *Q*, cele două sunt de fapt identice?' Ştim că această stare-senzaţie şi această stare cerebrală au loc în paralel, oare 'starea-senzaţie este identică cu starea cerebrală'? Cu cât discuţia merge mai departe pe acest făgaş, cu atât conceptul de *corelaţie* pare mai neproblematic. Intenţionez să arăt că, de fapt, corelaţia este totuşi problematică, nu în sensul că non-corelaţia este evidentă, ci în sensul epistemologic că *dacă* există o corelaţie nu putem şti care este. Problema nu depinde de asumarea materialismului, ci de faptul dacă credem că există cel puţin o corelaţie.

Teoria identităţii şi aprioricul

Interesul pentru teoria identităţii şi pentru alte forme ale monismului a fost reînviat, dacă nu de la început (odată cu Smart şi alţi teoreticieni ai identităţii), atunci, cel puţin din jurul anilor 1960, pentru motivul că filosofii o

considerau falsă, deşi nu se bazau pe o evidenţă empirică (ce fel de evidenţă empirică ar putea arăta că o stare-senzaţie nu este o stare cerebrală?) ci pe una *a priori*. Cine crede a priori că o stare-senzaţie nu este identică cu o stare cerebrală sau cine consideră că o asemenea identificare este *lipsită de sens* o face în acelaşi fel în care se consideră că numărul trei nu poate fi albastru. Prior, în anii 1950, sau continuatorii săi din anii 1960 considerau că ei, pur şi simplu, *ştiu* că stările-senzaţii nu pot avea natură fizică. Oponenţii ştiau, la rândul lor, că aceştia se înşeală. Dar o argumentare de o parte sau alta era imposibilă. Majoritatea ar spune 'Într-adevăr, nu se poate dovedi că o stare-senzaţie este o stare neurofiziologică, aşa cum nu putem dovedi că orice număr are un succesor sau nu putem dovedi că numărul trei nu este albastru, dar aceste lucruri sunt *cunoscute*; ele sunt *adevăruri ale raţiunii*. Ştim că este un nonsens sau o imposibilitate ca o stare-senzaţie să fie o stare neurofiziologică tot aşa de clar cum cunoaştem şi alte lucruri. Pe de o parte stă majoritatea, care ştie că stările-senzaţii nu pot fi stări cerebrale, pe de altă parte este minoritatea care ştie că majoritatea se înşeală. Cunoaşterea fiecărei părţi este *a priori*. Nu exista nici un mijloc de a depăşi această situaţie.

În 1951 W. V. Quine a publicat un studiu cu titlul 'Două dogme ale empirismului'[4]. Din acel moment s-a declanşat o eroziune a încrederii în noţiunea de adevăr 'a priori'. Quine arată că multe cunoştinţe pe care le consideram a priori au fost reevaluate. De pildă, să ne oprim la următorul exemplu: dacă cineva i-ar fi sugerat lui

[4] 'Two Dogmas of Empiricism' a apărut pentru prima dată în *The Philosophical Review*, 1951. A fost retipărit în *From a Logical Point of View*. (New York, 1961).

Euclid că două linii *drepte* perpendiculare pe o a treia *dreaptă* s-ar putea *intersecta*, Euclid ar fi răspuns că în mod necesar aşa ceva nu se întâmplă. Conform teoriei fizice de astăzi, totuşi se întâmplă. Razele de lumină care trec prin apropierea Soarelui se comportă în acest fel, nu pentru că ar fi curbe, ci tocmai pentru că îşi continuă traiectoria dreaptă într-o lume neeuclidiană.

O dată acceptată această situaţie, se ridică întrebarea 'Ce rămâne *a priori*?' (Quine a arătat convingător că tipul de aprioric admis de empirismul standard – bunăoară, noţiunea 'adevărat prin convenţie' – este incoerent, dar nu voi relua argumentele sale).

Cred, însă, că Quine a mers prea departe. Afirmaţia lui Quine că 'nici o judecată nu este imună la schimbare' sugerează că pentru orice judecată există circumstanţe sub care ar fi raţional să fie respinsă. Dar aceasta este destul de limpede fals: sub ce circumstanţe ar putea fi respinsă raţional judecata 'unele judecăţi nu sunt adevărate', oare trebuie să acceptăm că 'orice judecată este adevărată'?[5]

Dar dacă Quine se ridică împotriva lui *a priori* în mod exagerat, următoarea regulă este corectă: noţiunile noastre de raţionalitate şi modificare raţională nu sunt fixate prin reguli imuabile nici nu se află înscrise în natura noastră transcendentală, cum credea Kant, pentru că ideea de natură transcendentală, care se vrea *noumenală*, neinfluenţată de aspectele istorice şi biologice ale omului, nu are sens. Dacă noţiunile noastre de raţionalitate şi revizuire raţională sunt produsul experienţei noastre limitate şi cad sub incidenţa biologicului, este de aşteptat ca şi principiile considerate *a priori* sau 'conceptuale' să fie

[5] Discut atacul lui Quine asupra noţiunii de *a priori* în lucrarea 'Analiticity and Apriority: Beyond Wittgenstein and Quine', *Midwest Studies in Philosophy*, vol. IV, 1979 (Minnesota).

modificate din când în când potrivit vreunei experienţe neaşteptate sau vreunei inovaţii teoretice. O asemenea revizuire nu poate fi nelimitată: altfel nu am mai avea vreun concept de *raţionalitate*; dar, de regulă, limitele nu pot fi stabilite. Pe lângă situaţiile triviale (de felul 'Unele judecăţi nu sunt adevărate') nu putem avea siguranţa că vreo judecată este raţională, adică este adevărată în orice context (respectiv, într-un context dat), sau că este un adevăr 'necesar'. În general, trebuie să admitem pe considerente de simplitate, de utilitate şi plauzibilitate că poate exista un dat a priori. *Filosofia a devenit antiaprioristă.* Dar, odată ce trebuie să acceptăm despre cele mai multe adevăruri despre care credeam că sunt a priori că au un caracter relativ şi contextual, *trebuie să aducem un argument potrivit împotriva identităţii minte-corp.* Adepţii teoriei identităţii au reuşit să preîntâmpine o asemenea tentativă. Astfel, situaţia s-a schimbat.

Am folosit noţiunea de 'proprietate'; dar mi se pare că există cel puţin două noţiuni de proprietate care au rămas confundate[6]. Pe de o parte, este o noţiune mai veche pentru care este folosit termenul 'predicat' (de exemplu, în faimoasa întrebare 'existenţa este un predicat?'), iar pe de altă parte, există noţiunea de 'proprietare fizică' sau 'mărime fundamentală' etc. Atunci când un filosof are în vedere vechea noţiune, discursul despre proprietăţi poate fi înlocuit cu un discurs despre concepte. Pentru un asemenea filosof proprietăţile nu pot fi identice decât dacă există un adevăr conceptual care susţine că sunt aceleaşi; în particular, proprietatea de a avea senzaţii de o anumită calitate nu poate fi identică cu proprietatea de a se afla

[6] Vezi 'On Properties', Cap. 19 din volumul meu *Mathematics, Matter and Method, Philosophical Papers*, vol I (Cambridge, 1975).

într-o anumită stare cerebrală, dacă predicatele corespunzătoare nu sunt *sinonime* (în sensul larg de 'echivalenţă analitică'), iar principiul de individuaţie pentru predicate este că a fi *P* este acelaşi lucru cu a fi *Q* tocmai în cazul în care 'este *P*' este sinonim cu 'este *Q*'.

Fie afirmaţia că temperatura *înseamnă* energia cinetică moleculară. Pe de o parte, aceasta este o judecată de identitate a proprietăţilor. Ea asertează că proprietatea de a avea o temperatură anume este *în realitate* (într-un anumit sens al 'realităţii') *aceeaşi proprietate* cu proprietatea de a avea o anumită energie moleculară; sau (mai general) că *mărimea fizică* a temperaturii este una şi aceeaşi cu mărimea fizică a energiei cinetice moleculare. Dacă '*x* are cutare temperatură' nu este *sinonim* cu '*x* are anume energie cinetică moleculară', chiar dacă 'anume' este valoarea energiei moleculare care corespunde pentru 'cutare' valoare a temperaturii, atunci ceea ce fizicianul înţelege prin 'mărime fizică' este altceva decât ceea ce filosoful înţelege prin 'predicat' sau 'concept'.

Pentru a fi mai clar, în timp ce pentru ca predicatele *P* şi *Q* să fie aceleaşi, se cere sinonimia expresiilor '*X* este *P*' şi '*X* este *Q*' aceasta nu este cerută pentru coincidenţa proprietăţilor. Proprietăţile, spre deosebire de predicate, pot fi 'identice sintetic'.

Dacă admitem ceva de genul *identitatea sintetică a proprietăţilor*, atunci de ce să nu fie posibil ca proprietatea de a avea o anumită stare cerebrală să fie aceeaşi cu proprietatea de a avea o anumită senzaţie (în mod asemănător cu perspectiva lui Spinoza) – deşi nu sunt legate printr-un adevăr conceptual, ba chiar, s-ar părea că este *a priori* fals? Acesta este argumentul. Pe scurt, pe fondul antiaprioriismului, s-a formulat o nouă variantă a *proprietăţilor sintetic identice* şi, pe această bază, teoria identităţii şi, în particular, funcţionalismul, sunt relansate.

Acum, să luăm în considerare ce se întâmplă când la cele două aspecte se adaugă al treilea. Ce se întâmplă cu filosofii care: (1) sunt naturalişti antiapriorişti; (2) admit identitatea sintetică a proprietăţilor şi (3) susţin o perspectivă realistă tare a adevărului? Doresc să arăt că un asemenea filosof se confruntă cu probleme epistemologice serioase.

Creiere secţionate

Să ne oprim la un experiment pe care neurologii l-au efectuat acum 20 de ani, şi anume la faimosul experiment al 'creierului împărţit' sau al disocierii senzaţiilor. Vreau să investigăm importanţa acestui tip de experiment pentru teoria identităţii şi pentru noţiunea de *corelaţie*.

Potrivit modelului creierului ca sistem cognitiv analog unui computer, creierul are un limbaj, un limbaj intern (care poate fi strict intern sau poate fi un amestec de limbaj intern sau sistem de reprezentare şi limbaj public). Unii filosofi au inventat un nume pentru această limbă cerebrală ipotetică, 'mentaleză'. Să cercetăm ce se întâmplă când cineva are o senzaţie vizuală conform unui asemenea model (voi face apel la cunoştinţele mele de neurologie, chiar dacă nu sunt suficiente, dar nu cunosc pe nimeni care să ştie destulă neurologie). Există un singur scenariu posibil.

Când subiectul are o senzaţie, face o judecată; creierul 'înregistrează' ceva de tipul 'roşu la ora 12'. Astfel, calitatea (să-i spunem *Q*) corespunde, între altele, cu o *înregistrare în mentaleză*. De asemenea, este excitată intrarea în procesorul verbal central, centrul care este conectat cu sistemul vocal şi care dă seama de abilitatea creierului de a traduce în limbajul

public 'roşu acum'. Se poate ca judecata în mentaleză să fi fost transmisă dintr-un loc în altul înainte ca intrarea în centrul vorbirii să fie excitată. De asemenea, se produc diverse evenimente în zona vederii din cortex (care au fost studiate de neurologii Hubel şi Wiesel) care, după cum îmi imaginez, contribuie la înregistrările făcute în mentaleză şi la procesul vorbirii. Aceste 'înregistrări', 'intrări' şi alte evenimente pot avea loc în diferiţi lobi cerebrali: când *corpul calos* este secţionat, lobul drept (lob care nu este responsabil cu vorbirea) poate vedea roşu (sau cel puţin va răspunde afirmativ la stimuli care afectează numai acest lob), dar dacă cineva întreabă subiectul ce culoare are cartonaşul, acesta va răspunde 'nu văd cartonaşul'. Pe de altă parte, există un loc unde se formează memoria (respectiv, memoria pe termen scurt şi cea pe termen lung). Aproape sigur, nu este vorba de un lanţ cauzal linear; probabil că există ramificaţii şi cicluri, o întreagă reţea cauzală.

Dacă teoria identităţii este corectă, atunci starea-senzaţie *Q* este identică cu o stare sau alta a creierului. Un realist metafizic nu poate considera o asemenea relaţie drept urmare a unei convenţii sau a unei decizii, ori ca având o componentă convenţională *prin care* o stare cerebrală este identică cu *Q*. Ar trebui să trăim într-o lume în care, senzaţiile cu anumite calităţi pe care le resimţim sunt în mod real *unul şi acelaşi lucru* cu proprietăţi pe care le constatăm pe alte căi drept proprietăţile fizice ale evenimentelor cerebrale. (Mai bine spus, în care proprietatea de a avea o senzaţie de o anume calitate este tocmai proprietatea de a fi într-o stare cerebrală dată.)

Să ne oprim pentru moment asupra celor susţinute de această doctrină. Să presupunem că investigăm calitatea subiectivă de *roşu* (produsă, să spunem, prin activarea şi retragerea unui disc verde, ca imagine remanentă). Să presupunem că atunci când experimentez acest *roşu*, starea-senzaţie în care

sunt este identică cu o disjuncţie de stări ale creierului. Ea nu poate fi identică cu o stare unică maximală a creierului deoarece ştim că, dacă se îndepărtează vreun neuron, experienţa poate avea loc. Dar proprietatea ar putea fi disjunctivă, să spunem (*neplauzibil*), *dacă sunt excitaţi neuronii din zona bla-bla-bla cărora le corespund numere pare sau dacă sunt excitaţi cei cărora le corespund numere prime.* De fapt, ar fi posibilă o disjuncţie şi mai amplă. Am avea de-a face cu o serie imensă de stări neurologice a căror disjuncţie ar fi experienţa *roşu.*

Să mergem mai departe. Dacă sunt excitaţi neuronii pari din zona bla-bla-bla, am experienţa *roşu.* Dacă cerebroscopul spune 'nu, neuronii cu numere prime din zona bla-bla-bla sunt excitaţi', am experienţa *roşu.* Adică, nu pot spune în care dintre aceste două stări mă aflu. Dacă am experienţa *roşu* trebuie să fiu în una dintre ele, dar nu pot spune în care. Proprietatea *neuronii pari din zona bla-bla-bla sunt excitaţi* nu este observabilă. Chiar dacă am şti că teoria identităţii este adevărată nu pot spune, pornind de la senzaţii, care este proprietatea. Fie această proprietate P_1 şi fie proprietatea când *neuronii impari sunt excitaţi*, P_2. Starea-senzaţie este identică cu disjuncţia (P_1 sau P_2), care este, desigur, o a treia proprietate. P_1 şi P_2 nu sunt stări-senzaţie, ci numai disjuncţia lor ar fi o asemenea stare. Cu alte cuvinte, în această ontologie, *disjuncţia* a două proprietăţi care sunt ele însele neobservabile poate fi *observabilă.* Ajungem la funcţii logice complicate de proprietăţi neobservabile pentru a da seama de o observaţie simplă.

S-ar putea ca argumentarea de mai sus să sune ciudat. Un prieten al meu a făcut remarca: 'Să imaginăm un dispozitiv care poate detecta mionii fără a deosebi mionii de antimioni. Atât mionul, cât şi antimionul nu sunt observabili, dar disjuncţia lor este. Acest lucru poate părea paradoxal doar pentru cei care vor ca observaţia să fie mai

puţin pragmatică decât este'. Scopul meu nu este de a diminua importanţa unei perspective care este un program de cercetare foarte important în neuropsihologie, ci de a clarifica angajamentele sale ontologice. Dificultăţile, cum voi argumenta, nu apar din teoria identităţii ca atare, ci din conjuncţia între teoria identităţii şi realismul metafizic, conjuncţie pe care am numit-o perspectiva 'externalistă' asupra adevărului.

S-ar putea evita asemenea angajamente. De pildă, Carnap a spus (cel puţin într-o anumită perioadă) că discursul despre obiecte fizice derivă din discursul despre senzaţii şi că decizia că o stare cerebrală particulară este identică cu starea-senzaţie *Q* este o decizie de modificare a limbajului în care vorbim despre proprietăţile fizice, de a schimba concepţiile privind acele proprietăţi fizice.

Dacă discursul despre obiectele şi proprietăţile fizice este derivat din discursul despre senzaţii, am putea schimba regulile. Dar acest punct de vedere nu aparţine realismului metafizic cel puţin cu privire la obiectele şi proprietăţile fizice. Unii cred că pot fi realişti metafizici în ce priveşte *senzaţiile,* dar să nu fie realişti metafizici cu privire la *obiectele materiale* şi, dacă privesc obiectele materiale ca pe ceva maleabil, aceştia ar putea adopta teoria identităţii spunând 'sunt de acord cu acest tip de *convenţie* ca mijloc de a stipula sensuri'. Dacă sensurile nu ar fi fixate anterior, dacă ar reprezenta o reţea deschisă, nu ar apare nici o problemă cu privire la 'cum putem şti că o stare-senzaţie este identică cu o proprietate şi nu cu alta?'. Dacă această proprietate este vagă, atunci cu un simplu postulat, identitatea ar deveni o problemă de precizare a sensului. Dar aici îi am în vedere pe cei care cred că există o lume materială care nu este doar derivată din senzaţii; pe cei care cred că există proprietăţi fizice şi pe cei care susţin că

expresii precum 'neuronii de pe cutare sau cutare traiect sunt excitaţi' stau pentru proprietăţi fizice determinate, iar aceste proprietăţi fie *sunt* fie *nu sunt* identice cu stările-senzaţii.

De asemenea, filosofii care, precum Daniel Dennett, consideră că discursul cu privire la senzaţii este vag, care nu cred că există proprietăţi subiective bine definite, de pildă a fi într-o stare-senzaţie sau a avea senzaţii de un anumit tip, ar putea adopta o teorie a identităţii ca precizare a sensului, fixând, de data aceasta, nu sensul termenilor relativi la obiecte fizice, ci sensul termenilor psihologici. Dar, iarăşi, nu am avea de-a face cu un realism metafizic autentic.

Consider că un realist metafizic autentic este cel care crede 'da ştiu că această proprietate psihologică există (stare-senzaţie). Am avut-o. Pot să o recunosc. Cred că este o proprietate psihologică definită la care pot să mă refer. Ştiu că P_1 şi P_2 există, aşa încât există şi 'P_1 sau P_2', cât şi că o stare-senzaţie este identică cu aceasta sau nu este'. În acelaşi fel în care un fizician naiv ar putea spune 'aici nu este prezentă nici o convenţie' (cred că ar fi pe un drum greşit); 'nu există nici o convenţie în decizia că *temperatura este energia cinetică moleculară*, aşa că temperatura fie este fie nu este aceeaşi proprietate cu energia cinetică moleculară'. Aceasta este perspectiva de care mă voi ocupa.

Problema care apare este că, dacă cineva aderă la punctul de vedere al realismului metafizic, atunci sunt mai multe variante de luat în considerare. Posibilitatea care îmi vine mai întâi în minte este că o stare-senzaţie este identică cu proprietatea de a avea loc anumite evenimente în cortexul vizual *şi* a avea loc o 'înregistrare' în 'mentaleză' *şi* a avea o intrare în unitatea centrală a vorbirii *şi* a se activa mecanismul memoriei – starea senzaţie este gândită ca o

conjuncţie a acestor proprietăţi. Dar dacă invocăm posibilitatea disocierii, conjuncţia în întregul ei nu mai apare necesară. Poate că senzaţia este un eveniment simplu în cortex? (proprietatea de a avea o senzaţie este proprietatea 'reală' de a avea loc un eveniment în cortexul vizual).

Să explorăm, pentru moment, această variantă. Să presupunem că putem lăsa deoparte procesul care produce înregistrarea în mentaleză, sau intrarea în centrul vorbirii. Să ne imaginăm că am arătat subiectului un carton roşu în partea stângă a câmpului său vizual (aşa încât cartonul este vizibil numai pentru lobul drept, cum spun neurologii). Evenimentul corespunzător din cortexul vizual se va petrece în lobul drept, dar dacă subiectul este întrebat: 'Vezi ceva roşu?', el va răspunde 'Nu'.

Acum, prin criteriul pe care l-am folosit pentru a decide dacă cineva are sau nu o senzaţie, respectiv criteriul sincerităţii, trebuie să acceptăm că subiectul *nu are senzaţia de roşu*, iar teoria că *Q* (caracterul calitativ) este identic cu evenimentul din cortex trebuie respinsă. Dar, s-ar putea obiecta: 'Nu aţi reuşit să respingeţi teoria în nici un fel. Ce fel de observator este acesta? Creierul lui este secţionat în două.' Pentru un observator normal, *Q* este identic cu proprietăţile cortexului vizual. Observatorii care nu se află într-o stare normală nu intră în discuţie.

Dificultatea persistă, deoarece există teorii ale identităţii care nu pot fi *deosebite observaţional*[7], adică ele

[7] Noţiunea 'imposibil de deosebit observaţional' a fost introdusă în lucrările cu privire la teoria spaţio-temporală a lui Clark Glymour şi David Malament în *Foundations of Space-Time Theories*, Earman, Glymour şi Stachel (eds.), *Minnesota Studies in the Philosophy of Science*, vol. VIII (Minnesota University, 1977). Problema analogă din această teorie este existenţa unor spaţio-temporalităţi posibile (cum ar fi spaţio-temporalitatea

conduc la aceleaşi predicţii cu privire la experienţa tuturor *observatorilor aflaţi în condiţii normale.*

Să luăm în considerare ipoteza când cineva nu are senzaţia de roşu decât dacă este activată intrarea în centrul vorbirii. Cum ar putea fi confirmată sau respinsă? Am putea crede că o posibilitate în acest sens ar fi dacă secţionăm *corpul calos* şi constatăm că există memorie fără ca centrul vorbirii să fie activat; respectiv, am întreba subiectul dacă are senzaţia de roşu. El spune 'Nu'. După aceea refacem unitatea corpului calos şi întrebăm: 'Ai avut senzaţia de roşu?' Subiectul ar putea răspunde: 'Da, dar ceva era în neregulă, aveam senzaţia de roşu când m-ai întrebat dar m-am auzit, sincer fiind, spunând că nu o am'. (Am spus că este normal pentru subiecţi să 'reconcilieze' sau să raţionalizeze situaţiile de acest tip considerându-le imaginare). Oare, un asemenea raport arată că exista o senzaţie de roşu în absenţa excitării centrului vorbirii?

Nu. Dacă Daniel Dennett (care, la un moment dat, a susţinut că senzaţia *este* excitarea centrului vorbirii sau ceva apropiat[8]) ar dori să reconcilieze raportul subiectului cu teoria sa, ar trebui să spună că 'Nu neg că ulterior a avut loc evenimentul psihologic de *reamintire a unei senzaţii anterioare.* Neg că *senzaţia* a avut loc anterior'. După

---

permisă de teoria relativităţii) care diferă în ceea ce priveşte proprietăţile topologice globale, dar în care observatorii ar avea exact aceleaşi experienţe. Asemenea consideraţii sunt respinse de obicei pornind de la 'consideraţii de simplitate' potrivit cărora este suficientă o singură spaţio-temporalitate – cea în care trăim; problema care apare în acest caz (cum subliniază Malament) este că teoria fizică (relativitatea generală) nu spune că trăim în *cea mai simplă* spaţio-temporalitate compatibilă cu legile sale.

[8] Modelul Dennett al conştientizării este prezentat în studiul 'Towards a Cognitive Theory of Consciouness', retipărit în *Brainstorms* (Bradford Books, 1978).

cealaltă teorie, subiectul are mai târziu experienţa reamintirii unei senzaţii pe care a avut-o mai devreme.

Dezacordul este unul de fapt. Cei mai mulţi neurologi cred că, în cazul pacientului cu creierul secţionat, lobul drept este cel al conştiinţei. Prin urmare, sunt tentaţi să spună că există o senzaţie de roşu cu toate că centrul vorbirii nu este afectat. ('Există două locuri ale conştiinţei' sa afirmă frecvent). Cel puţin unul dintre neurologi, Eccles, susţine că lobul drept luat separat (sau lobul stâng pentru pacienţii care au sediul vorbirii în lobul drept) nu este *conştient*. Pentru Eccles, conştiinţa cere *unitatea* cerebrală; separarea lobului stâng lasă loc unui comportament conştient simulat care nu arată că este 'locul conştiinţei'.

Nici apelul la maxime metodologice, cum ar fi 'alege teoria cea mai simplă' nu este de ajutor, deoarece nu suntem în faţa unui caz relevant de simplitate de partea perspectivei integraliste sau a celei separatiste în ce priveşte conştiinţa. Poate cea separatistă să fie mai simplă sub un aspect; ea spune că anumite capacităţi comportamentale (pe care le posedă lobul drept, chiar dacă nu este posibilă vorbirea) sunt suficiente pentru prezenţa conştiinţei, iar acest lucru este conform cu observaţia că animalele (care nu au abilităţi lingvistice) sunt conştiente. Dar există multe diferenţe între un animal care are un creier intact, cu procese cerebrale complete chiar dacă nu poate vorbi, şi unul cu creierul secţionat. Dacă o asemenea situaţie nu ar fi atât de apropiată nouă, dacă nu am avea tendinţe puternice înspre realismul metafizic privind *senzaţiile*, nu ar trebui ca, în acord cu cele mai bune intuiţii metodologice, să-l privim drept un caz care trebuie *legiferat* mai degrabă decât impus?

Pe scurt, există un număr de teorii ale identităţii care nu pot fi deosebite observaţional. Dacă teoreticienii identităţii au dreptate nu ar exista nici o posibilitate de a

spune *în ce fel au dreptate*: nu este posibilă stabilirea stării cerebrale concrete cu care o stare-senzaţie este identică. Aspectul este suficient de important pentru a-l dezvolta în cele ce urmează.

Thomas Nagel[9] a făcut afirmaţia plauzibilă că nu putem să ne dăm seama cum este să fii liliac. Dar *de ce* o asemenea afirmaţie ar trebui să fie plauzibilă? Cu câţiva ani în urmă am citit o carte interesantă a lui Donald Griffin. Mi-am dat seama că liliecii nu sunt prea deosebiţi de celelalte mamifere. Cei mai mulţi dintre noi credem că ne-am putea imagina senzaţiile pe care le are un câine sau o pisică. Ce dificultate apare în legătură cu liliecii?

Ei bine, liliecii pot auzi sunete cu câteva octave mai înalte decât noi. Nu pot să-mi imaginez ce înseamnă să fii liliac în sensul că nu-mi pot da seama cum arată senzaţiile dobândite prin sonar. Dar un asemenea efort de imaginaţie este, oare, atât de dificil? În tinereţe suntem în stare să auzim sunete cu o octavă mai înalte decât la maturitate, dar tabloul subiectiv nu se modifică: cele mai înalte sunete pe care le aud acum pot fi cu o octavă mai jos, dar ele au aceeaşi ascuţime, fiind cele mai înalte sunete pe care le pot auzi. Poate aşa se întâmplă şi cu sunetele cu cinci octave mai înalte: sunt aidoma unui chiţăit scurt şi înalt.

Să ne imaginăm o dezbatere între doi filosofi sau psihologi, unul dintre ei afirmă că *quale** pentru un liliac nu este deloc precum *quale* uman. *Qualia* liliacului sunt inimaginabil de diferite de *qualia* umane. Niciodată nu am fi în stare să ne imaginăm cum e să fii liliac (sau chiar câine ori pisică). Celălalt filosof, vă puteţi imagina, răspunde: 'Prostii! Poate există unele senzaţii ale liliacului pe care nu

[9] 'What is it to be a bat?', retipărit în N. Block, *op.cit.*

* (componenta calitativă a unei senzaţii – *n. trad.*)

le pot avea. Există şi senzaţii ale altor oameni (de pildă, senzaţiile femeilor) pe care probabil că nu le pot avea, dar de aici nu pot considera spaţiul psihologic al altor oameni ca inimaginabil pentru mine. De ce ar trebui să cred că spaţiul vizual al liliacului, de pildă, ar fi cu totul diferit de spaţiul meu vizual? (*N.B.* liliecii văd foarte bine, în ciuda credinţelor populare). Cu câteva adaptări care ţin de optica ochiului liliacului sau de auzul său, auzul liliacului este ca şi auzul meu, precum durerea liliacului este ca şi durerea mea.' Cum am putea decide între aceste două poziţii?

Deoarece numărul şi configuraţia neuronilor sunt diferite (centrul auditiv la liliac poate ocupa 7/8 din creier), proprietăţile la nivelul neurologic complet – numărul neuronilor excitaţi – care este identic cu o *quale* în cazul unui liliac în cazul când teoria identităţii este corectă, nu poate fi aceeaşi proprietate cu care este identică acea *quale* în cazul omului. Sau poate fi? Să presupunem că, atunci când liliacul are o anumită senzaţie vizuală (produsă prin vederea unor obiecte roşii), acesta are proprietatea disjunctivă (P_1 sau P_2), unde P_1 şi P_2 sunt stări specifice maximale ale creierului liliacului. (De fapt, este o proprietate disjunctivă mult mai complicată cu mii de termeni, dar o simplificăm). În plus, atunci când eu am o anumită senzaţie vizuală (produsă de vederea unor obiecte roşii) creierul meu are proprietatea disjunctivă (P'_1 sau P'_2). Să considerăm următoarele două teorii: (1) componenta calitativă a senzaţiei liliacului (să-i spunem *roşuL*) este identică cu (*sau cel puţin corespunzătoare cu*) proprietatea disjunctivă (P_1 sau P_2) iar componenta calitativă a senzaţiei umane de roşu (*roşuH*) este identică cu (*sau cel puţin corespunzătoare cu*) proprietatea diferită (P'_1 sau P'_2); (2) componenta calitativă a senzaţiei liliacului este identică cu componenta calitativă a senzaţiei mele (*roşuL=roşuH*) şi ambele sunt identice cu (*sau cel puţin*

corespunzătoare cu) proprietatea disjunctivă mai complexă (P_1 sau P_2 sau P'_1 sau P'_2).

Potrivit primei teorii, liliacul are experienţe diferite în vreme ce, pentru a doua teorie, are aceleaşi experienţe; dar aceste două teorii conduc la aceleaşi predicţii privind ceea ce simt observatorii umani.

Oare maxima metodologică 'alege teoria mai simplă' este de ajutor? Nu este limpede cum ar putea fi. Ned Block a arătat că prima teorie este mai simplă sub un aspect (*quale* este identificată cu o proprietate fizică mai simplă) dar a doua este mai simplă sub un alt aspect (nu este 'şovină'; ea permite ca doi subiecţi care nu au aceeaşi constituţie fizică să aibă aceleaşi *qualia*). Iată că lipsesc principiile care să permită alegerea unei teorii sau a alteia. Dar pentru ce ar trebui să existe asemenea principii? De ce să nu abandonăm, aşa cum a argumentat Wittgenstein, realismul metafizic cu privire la senzaţii şi cu privire la 'aceleaşi' (ca aplicat la senzaţii) şi să le tratăm ca un caz de legiferat mai curând decât unul descoperit?

La sfârşit mă voi ocupa cu trei teorii despre care sunt sigur că sunt false, dar prin care este dificil sau imposibil să se arate dacă realismul metafizic este corect. Acestea sunt: (1) *roşuH* este identic cu o stare funcţională (sau quasi-funcţională), respectiv starea de a fi în orice stare materială care a jucat în viaţa noastră rolul de a semnala prezenţa roşului obiectiv; (2) pietrele acestea au *qualia* (în interiorul lor au loc evenimente calitativ similare cu senzaţiile noastre vizuale); (3) naţiunile au conştiinţă.

Să ne oprim la (1). Ne reamintim argumentul pe care l-am folosit pentru a arăta că *roşuH* nu poate fi o stare funcţională: dacă identificăm *roşuH* cu o stare funcţională, respectiv de *a fi în orice stare materială* (*cerebrală*) *care semnalează prezenţa roşului*, atunci nu s-ar explica experimentul inversării spectrului (cel puţin în situaţia de

amnezie), dacă mă aflu în acea stare funcţională când văd ceva roşu atât înainte cât şi după inversiunea spectrului (între timp având loc schimbările lingvistice corespunzătoare şi, dacă este necesar, apariţia unui atac de amnezie). De pe poziţiile realismului metafizic, este posibil să sufăr o inversiune spectrală (cu toate că memoria este afectată de atacul de amnezie). Cazul este mai edificator dacă nu presupun atacul de amnezie şi îmi amintesc că spectrul a fost inversat; dacă ajustările lingvistice au loc simultan, există un sens în care ce era înainte 'senzaţia de verde' joacă rolul funcţional de 'a semnala prezenţa *roşului* obiectiv în mediu'.

Acest argument arată numai că *roşuH* nu este identic cu starea funcţională menţionată. Nu arată că nu este identic cu starea funcţională mai complicată *de a fi în orice stare materială care anterior a avut funcţia respectivă.* S-ar putea obiecta că aceasta este o proprietate ciudată, dată printr-o funcţie logică greoaie de stări funcţionale. Dar de ce să fie mai complicată o asemenea funcţie logică de stări funcţionale decât o disjuncţie de proprietăţi fizice? Oare lumea preferă disjuncţiile de proprietăţi fizice conjuncţiilor de proprietăţi funcţionale?

Să considerăm teoria (2). Fie P_3 proprietatea de a fi piatră şi să admitem ipoteza că *roşuH* este identic cu proprietatea disjunctivă (P_1 sau P_2 sau P'_1 sau P'_2 sau P_3). De bună seamă că şi pietrele au o asemenea proprietate. Astfel, sub această ipoteză, evenimentul de calitatea *roşuH* are loc şi în cazul pietrelor. (Acestea nu simt roşul în sensul funcţional al simţirii roşului, dar un eveniment cu caracter calitativ al evenimentului care joacă rolul funcţional de a simţi roşu pentru noi are loc şi în cazul pietrelor.) Sau să admitem ipoteza mai complexă că pietrele au *qualia* diferite la momente diferite. Sau chiar ipoteza că una dintre aceste

presupuneri (fără a stabili care dintre ele) este corectă. Am putea spune 'asemenea ipoteze sunt nebuneşti'. Într-adevăr, sunt, dar fiecare dintre ele generează o teorie sănătoasă. Nici una nu poate fi exclusă pe baza observaţiei sau experimentului, deoarece sunt indistincte observaţional în raport cu perspectiva standard.

Am putea crede că asupra acestor teorii s-ar putea decide printr-un principiu metodologic precum *să nu se atribuie nici o proprietate unui obiect fără motiv*. Desigur, acest principiu nu conduce la falsitatea unei teorii (uneori lucrurile pentru care nu există motive să le credem, se pot dovedi adevărate), dar cel puţin ne spune dacă avem *justificare* să considerăm o teorie falsă. Dar, oare, nu există nici un motiv pentru a susţine vreuna dintre aceste teorii (teza că o asemenea teorie este corectă, iar pietrele au *qualia*)? Oare din argumentul că, dacă eu am *qualia* şi fizicalismul este adevărat (mulţi filosofi cred că există motive suficiente pentru a accepta fizicalismul) atunci există cel puţin un obiect fizic în care au loc evenimente cu un caracter calitativ, nu ar rezulta că asemenea evenimente pot să aibă loc în *toate* obiectele fizice? Dacă am putea arăta că există ceva legat de *quale* care *cere* un anumit 'rol' funcţional care este specific oamenilor, acest argument este respins; dar tocmai asta ne spun cei care cred în *qualia* ca obiecte reale metafizice: spun că un asemenea element specific uman nu se poate determina.

În final să ne oprim asupra lui (3). Să admitem ipoteza că durerea sau teama este o stare funcţională care poate fi manifestată de organisme sau naţiuni. Cu alte cuvinte, să presupunem că atunci când Statele Unite anunţă 'Statele Unite se tem de...' chiar aşa se întâmplă. Desigur, asta nu putem să o ştim. Poate că cititorul crede despre un grup că se poate manifesta de parcă poate simţi ceva

asemănător cu durerea; dar nu crede că Statele Unite ar simţi într-adevăr durere. Potrivit ipotezei (3), cititorul s-ar înşela: Spiritul naţional ar simţi realmente durerea.

Această ipoteză este în legătură cu o discuţie interesantă între filosofii mentalului. Funcţionaliştii (între care mă număr şi eu) ar putea invoca următorul argument anti-şovin: în principiu, diferenţa dintre un robot şi un om (în ce priveşte organizarea funcţională) ar putea fi redusă la detalii legate de alcătuirea fizică sau chimică. Am putea imagina un robot asemenea nouă până la nivelul reţelelor neuronale. (Ar putea avea carne şi sânge, dar nu creier). Singura diferenţă ar fi că, în timp ce noi avem neuroni alcătuiţi din carbon, hidrogen, proteine etc. robotul ar avea neuroni electronici, dar legăturile interneuronale ar corespunde. Acum, dacă nu suntem 'şovini carbon-hidrogen' şi să credem că hidrogenul şi carbonul sunt mai conştiente, de ce să nu recunoaştem că robotul este o persoană care întâmplător are mai mult metal şi mai puţin carbon şi hidrogen?

Argumentul a provocat următoarea replică: 'ei bine, în locul acestor componente electronice, neuroni electronici legaţi împreună ca şi neuronii umani, să presupunem că ar fi oameni în miniatură, cercetaşe şi cercetaşi'. Nu este nevoie ca aceşti micuţi să ştie că fac parte dintr-o asemenea reţea sau să vadă altceva decât o incintă întunecoasă sau o mulţime de asemenea incinte în care fac schimb de semnale. (Timpul, pentru ei, ar trece foarte repede relativ la timpul nostru). I-am putea asimila muncitorilor exploataţi. 'Acum, ar continua replica, nu ai considera că acest sistem este "conştient" numai pentru că ştim că micuţii îşi mişcă propriul corp. O organizare funcţională (chiar ca a noastră) nu este suficientă pentru a justifica aplicarea unui predicat precum "conştient"'.

Replica la această replică era negaţia că 'robotul hidră cu multe capete' (cum a fost numit) are aceeaşi organizare funcţională pe care o avem şi noi. Dar există o replică mai radicală pe care am putea să o dezvoltăm. Am putea întreba: 'de ce nu ar trebui să considerăm conştient un asemenea robot? Dacă primul argument este corect (şi cred că este) dacă robotul cu creier pozitronic ar fi conştient, de ce ar trebui ca faptul că neuronii întregului robot ar fi mai conştienţi să însemne că întregul sistem este mai puţin conştient? În ultimă instanţă, suntem o societate alcătuită din mici animale. Celulele noastre sunt, luate individual, animale. S-ar putea ca ele să aibă senzaţii chiar rudimentare (cine ştie?) altele decât senzaţiile *noastre*.' Acum, dacă dezvoltăm această idee, odată ce acceptăm că robotul hidră este conştient (deşi neuronii săi sunt cercetaşi şi cercetaşe) atunci de ce nu ar fi astfel şi Statele Unite?

Nu afirm, de bună seamă, că Statele Unite au *aceeaşi* organizare funcţională cu *homo sapiens*. D bună seamă că nu. Dar există alte similarităţi. Statele Unite au organe defensive. Au organe digestive, hrănindu-se cu petrol sau cupru etc. Au o funcţie excretoare (poluantă). Oare nu avem de-a face cu organizarea unui mamifer căruia îi putem atribui durere?

Cât de bine este definită 'componenta calitativă'?

Până acum nu ne-am întrebat dacă este suficient de clar ce înseamnă ca două senzaţii să aibă aceeaşi 'componentă calitativă', chiar dacă la nivel introspectiv nu este cazul. Problema are sens pentru motivul că a fi acelaşi depinde de conceptualizările anterioare, cum se vede din faptul că

spunem despre tăblia mesei că este rotundă chiar când o vedem sub un anumit unghi.

În cazul tăbliei mesei, filosofii şi psihologii au argumentat încă din secolul al XIX-lea că, un subiect care are ca 'date ale simţurilor' elipse, *crede* numai din motive optice că lucrurile sunt rotunde (cu excepţia introspecţioniştilor consecvenţi) sau, că are un 'gestalt' rotund deşi datele înfăţişează elipse. Multe dintre experienţe sunt adaptate descrierilor. Nici neurologia nu a rezolvat această dispută: imaginea eliptică pe retină produce fără îndoială evenimente corespunzătoare în creier şi dacă le identificăm cu 'senzaţiile vizuale' atunci ajungem la ceva asemănător cu povestea clasică 'date ale simţurilor eliptice plus inferenţe inconştiente'; judecarea caracterului experienţei ('văd o tăblie de masă rotundă') corespunde 'înregistrărilor' şi 'intrărilor' în creier, iar dacă le identificăm *pe acestea* cu senzaţia vizuală, am putea, la fel de bine, adopta povestea că nu există date al simţurilor eliptice decât dacă judecăm despre ceva că arată eliptic. De ce nu am putea accepta ambele versiuni? După cum spune Goodman despre subiectul pus să descrie o mişcare aparentă,

> Cel mai bun lucru pe care îl putem face este să specificăm prin clasificarea termenilor, vocabularul pe care îl foloseşte subiectul, cerându-i să descrie ce vede în termeni perceptuali sau fenomenali mai degrabă decât fizici. Indiferent dacă se ajunge la răspunsuri diferite, se aruncă o lumină nouă asupra întregului proces. Trebuie evidenţiat faptul că aparatul observaţional este modelat de faptele observate, dar fără a identifica fizicul cu realul şi perceptualul cu aparentul. Perceptualul nu este o versiune a faptelor fizice, aşa cum fizicul nu este o versiune artificială a faptelor perceptuale.[10]

[10] *Ways of Worldmaking*, pp. 92–3.

Dacă văd o faţă de masă la două momente diferite ale zilei am *aceeaşi* senzaţie de roşu? Sau am senzaţii diferite şi nu-mi dau seama de diferenţă (decât dacă aş fi pictor)?

Un caz aparte este cel al *adaptării*. Dacă unui subiect i se dau ochelari care inversează imaginea, după un timp, va vedea lucrurile din nou normal. Oare datele simţurilor s-au răsturnat? Sau avem de-a face cu date alterate ale simţurilor care au fost interpretate 'sus' sau 'jos'? Subiectul nu poate preciza momentul în care lucrurile revin la normal. (Cititorii care, ca şi mine, poartă ochelari se pot întreba: jumătatea de jos a câmpului vizual arată diferit chiar dacă nu ne dăm seama?) Pe de altă parte, există schimbări la care subiectul nu se poate acomoda (de fapt, ne putem obişnui numai cu schimbări destul de slabe) şi am în vedere inversarea culorilor, fenomenul acomodării, cu siguranţă, aruncă îndoieli asupra bunei definiri a noţiunii de 'aceeaşi componentă calitativă'.

Realismul despre *qualia*

Am trecut în revistă o serie de dificultăţi sceptice. Scopul lor nu este să arate că teoria identităţii este eronată (sau că teoria corelaţiei este greşită – totuşi pot fi înţelese ca dificultăţi cu care se confruntă ambele teorii), ci că, dacă sunt adevărate, atunci există un mare număr de moduri alternative prin care pot fi precizate detaliile, aşa încât nu se poate şti niciodată care dintre ele este adevărată. Necunoscând adevărul, nu ştim care este răspunsul la marile provocări tradiţionale, de pildă dacă pietrele sau alte obiecte neînsufleţite au sau nu *qualia*, dacă liliecii sau ale specii au *qualia* de acelaşi fel cu noi sau nu, dacă grupurile simt durerea etc.

De ce ar trebui să credem că este posibil logic ca o piatră să simtă durere (ca un eveniment cu un caracter calitativ asemănător cu durerea umană să se petreacă în piatră)? Poate că Russell ne dă unele sugestii privind acest tip de realism metafizic. Russell era realist atât în ce priveşte *qualia* cât şi în ce priveşte universaliile. Mai mult, el considera *qualia* drept universalele cele mai autentice. Un universal este, înainte de toate, un mod în care lucrurile se pot asemăna; conform lui Russell se pare că unele dintre cele mai primitive şi fundamentale exemple a modului în care lucrurile pot fi asemănătoare sunt asemănările calitative privind senzaţiile noastre. *Qualia*, pentru Russell, sunt universale *par excellence*.

Universalele sunt gândite de către realiştii tradiţionali ca bine definite: *cuvintele* pot fi vagi, dar universalele însele nu sunt. (Un cuvânt este vag când stă pentru o mulţime vagă de concepte, spunea odată Gödel; dar conceptele sunt bine definite).

Astfel, dacă *qualia* sunt universale şi universalele sunt bine definite, ele trebuie să fie bine definite dacă orice lucru sau eveniment dat – inclusiv în cazul creierului secţionat, al pietrelor sau al grupurilor şi al evenimentelor legate de acestea – se manifestă sau nu ca o anumită *quale*. Dacă *quale* este gândită independent de funcţia pe care o îndeplineşte, dacă se crede că este în întregime contingent că o asemenea componentă calitativă a ceva are o funcţie specifică, atunci apare ca o posibilitate logică pentru creierul secţionat sau pentru pietre să aibă acea *quale*.

Un filosof ca mine, care doreşte să nege că aceste posibilităţi au sens (deşi unii le acceptă – există tentaţia de a considera legitim ca lobul drept al creierului secţionat să fie considerat locul conştiinţei) trebuie să aibă grijă să clarifice că nu apără vreo formă de behaviorism. A spune despre

qualia că nu sunt entităţi bine definite nu este acelaşi lucru cu a nega existenţa lor, sau cu afirmaţia că sunt reductibile la comportament etc. Multe noţiuni, deşi sunt vagi, au referenţi determinaţi. Noţiunea de 'casă' nu este bine definită în cazul igluului (igluul este o casă?), al bordeiului, sau în alte cazuri. Dar de aici nu se poate trage concluzia că nu există case. Analog, din aceea că nu există nici un fapt care să decidă dacă lobul drept este cel al conştiinţei, nu rezultă că nu există fiinţe conştiente.

Similaritatea calitativă a senzaţiilor este definită oarecum prin extensie: dacă am o senzaţie de roşu după o senzaţie de verde ştiu că am senzaţii deosebite (şi ştiu aceasta fără a compara rolul lor funcţional), iar dacă am o senzaţie de roşu urmată de aceeaşi senzaţie de roşu ştiu (dincolo de vaguitatea la care m-am referit anterior) că am senzaţii similare. Dar, pentru un internalist, nu urmează că există de fiecare dată un fapt concret care determină două senzaţii (să lăsăm deoparte două evenimente oarecare) să fie similare sau nu.

Fie *E* evenimentul de a avea o senzaţie particulară la un moment dat şi *E'* un eveniment fizic oarecare dintr-o piatră. Sugestia că aspectul calitativ al lui *E* (să spunem *roşuH*) ar putea fi identic sau corelat cu o proprietate de tipul (P_1 sau P_2 sau P_3) (unde P_3 este proprietatea *a fi o piatră*) este nepotrivită unei sensibilităţi sănătoase. Sugestia că *E* şi *E'* ar putea fi asemenea este absurdă. Am văzut deja că ipoteza este absurdă pentru că violează maxima metodologică 'nu atribui *qualia* unui obiect fără raţiune'. Chiar dacă această explicaţie funcţionează, consecinţele ei depăşesc imposibilitatea ca o piatră să aibă *qualia* (sau incoerenţa noţiunii corespunzătoare). Dacă *aceasta* este tot ce este greşit cu ipoteza că pietrele au *qualia* atunci putem afirma: *este posibil din tot ce ştim ca pietrele să aibă qualia, dar este a priori improbabil să fie aşa.*

De fapt, ipoteza că pietrele au *qualia* este incoerentă într-o măsură mai mare decât ipoteza creierelor în

container; ca şi pentru ipoteza creierelor în container, presupune o teorie 'magică' a referinţei. Orice om întreg la minte consideră *E* şi *E'* atât de deosebite încât problema similarităţii calitative (în sensul în care două senzaţii pot fi similare calitativ, de pildă, când le simţim în acelaşi fel) nici nu se pune. Dar realistul metafizic, deşi nu neagă acest fapt, crede că *E* şi *E'* ar putea fi (posibilitate logică) similare în vreun fel, *deşi este un act nebunesc să credem asta.* El crede astfel deoarece trăieşte cu iluzia că, având senzaţia respectivă, cu o anumită componentă calitativă, cu o modalitate proprie în care este simţită, cu un rol funcţional, însoţită de gânduri şi judecăţi, forţează cumva expresia 'modalitatea în care senzaţia este simţită' (sau un înlocuitor tehnic, cum ar fi 'componenta calitativă al senzaţiei, sau *roşuH*, ori 'această *quale*') să se refere la un 'universal' definit, la o proprietate absolută bine definită a evenimentelor individuale. Dar nu acesta este cazul.

Dacă ar exista roboţi funcţionali izomorfi cu noi, cu care să lucrăm, argumentează ei, ne-am împrieteni cu unii dintre ei, am accepta că au conştiinţă. (Am avea totuşi îndoieli dacă ei ar avea *aceleaşi qualia* ca şi noi; dar despre aceşti roboţi nu ne-am pune problema dacă au aceleaşi *qualia* ca şi noi mai mult decât în cazul liliecilor sau câinilor). Să presupunem că avem de-a face cu roboţi hidră. (Să ne imaginăm că aceştia au *evoluat* printr-un proces biologic precum animalele pe Pământ). Ce am simţi cu privire la *ei*?

O vreme nu am fi siguri în legătură cu un asemenea caz bizar, s-ar părea că am putea ajunge să le atribuim conştiinţă (dacă suntem în legătură mai adesea cu întregul robot şi nu cu neuronii săi conştienţi – cercetaşii sau cercetaşele din povestea mea); dar probabil s-ar înregistra diferenţe de opinii. Dacă ajungem în situaţia de a fi siguri că roboţii hidră sunt conştienţi, atunci am putea începe

încet-încet să credem acelaşi lucru şi despre Statele Unite? Nu ştiu.

Perspectiva pe care am susţinut-o legat de toate aceste cazuri este că nu există nimic *ascuns*, nu există fapte noumenale despre entităţi reale conştiente sau despre entităţi reale non-conştiente sau despre calităţi care sunt în realitate similare sau diferite. Există numai fapte empirice: pietrele şi naţiunile sunt mult diferite de oameni şi animale; roboţii de orice factură sunt tipuri de obiecte intermediare etc. Pietrele şi naţiunile *nu sunt* conştiente; adică faptul că naţiunile nu au conştiinţă se petrece în mod efectiv.

Această teză depinde de standardele privind acceptabilitatea raţională, justificarea şi adevărul ultim, ca şi de standardele de similaritate care, la rândul lor, depind de moştenirea noastră biologică şi culturală (de pildă, dacă am interacţionat sau nu cu 'roboţi inteligenţi'). Dar ceva asemănător este adevărat mai mult cu privire la limbajul de fiecare zi, pentru cuvinte precum 'casă', 'persoană', 'zăpadă' şi 'maro', de exemplu. Un realist care acceptă această rezolvare a problemei despre *qualia* ar exprima-o spunând '*qualia* nu există în realitate', sau *qualia* aparţine 'clasei secunde a sistemului nostru conceptual'; dar ce reprezintă acea noţiune de existenţă care pune *casele* de aceeaşi parte cu *neexistentele*? Lumea noastră este una omenească, aşa încât ce este conştient şi ce nu este conştient, cine are senzaţii şi cine nu are, ce este similar calitativ şi ce este deosebit sunt dependente, în ultimă instanţă, de judecata umană privind asemănarea şi deosebirea.

Două concepţii despre raţionalitate

În capitolele anterioare am vorbit despre raţionalitate şi despre 'acceptabilitate raţională'. Dar raţionalitatea nu este uşor de abordat.

Problema are analogii cu alte domenii. Acum câţiva ani studiam comportamentul unui tip natural de cuvinte, de pildă *aur,* şi am ajuns la concluzia că extensiunea termenului nu este determinată numai de 'setul de reguli semantice' sau de alte norme instituţionalizate. Normele pot face ca anumite obiecte să fie *exemple paradigmatice* ale aurului; dar nu determină întreaga extensiune a termenului, nefiind imposibil ca un exemplu paradigmatic să nu fie de fapt aur, cum s-ar întâmpla dacă normele ar *defini* ce este aurul.

Suntem tentaţi să considerăm că un lucru aparţine

unui anumit tip chiar dacă testele *actuale* nu sunt suficiente pentru a arăta că este membru al tipului respectiv, dacă vreodată ar reieşi că are aceeaşi natură esenţială cu (sau mai vag este 'suficient de asemănător') exemplele paradigmatice (sau cu marea majoritate dintre ele). Ce anume este natura esenţială, sau ce înseamnă a fi suficient de similar depinde de tipurile naturale şi de context (ceaiul poate fi apă în unele contexte); dar pentru aur contează compoziţia finală încă de pe vremea când vechii greci au gândit-o ca determinând comportamentul substanţei. Numai acceptând că vechii greci înţelegeau prin *chrysos orice are aceeaşi natură esenţială* cu exemplele paradigmatice, vom înţelege căutările lor pentru noi metode de a detecta aurul contrafăcut (care au condus la testul lui Arhimede), ca şi speculaţiile lor fizice.

Este tentant să alegem aceeaşi cale şi în cazul raţionalităţii şi să spunem că o convingere este raţională nu datorită normelor de raţionalitate dintr-o cultură, ci datorită unei *teorii ideale* a raţionalităţii, o teorie care ar evidenţia condiţiile necesare şi suficiente pentru ca o convingere să fie raţională în circumstanţe relevante în orice lume posibilă. O asemenea teorie ar da seama de exemplele paradigmatice, aidoma unei teorii ideale despre aur; mai mult, ar putea să meargă dincolo de ele şi să producă criterii care ne-ar permite să înţelegem cazurile pe care nu le putem vedea deocamdată în profunzimea lor, aşa cum teoria actuală a aurului ne permite să înţelegem situaţii pe care chiar cei mai străluciţi greci antici nu reuşeau să le soluţioneze. O dificultate generală care apare în tratarea unor termeni precum 'raţional', 'rezonabil', 'justificat' etc. drept termeni pentru tipuri naturale constă în aceea că tentativele actuale de a stabili generalizări puternice pentru toate convingerile considerate raţionale

sunt prea sărace. Există legi stricte şi universale cărora li se supune aurul, care fac posibilă descrierea aurului drept ceea ce se supune acestor legi; dar care sunt şansele să găsim asemenea generalizări cărora să se supună toate situaţiile de convingeri justificate raţional?

Faptul că asemenea şanse sunt reduse nu înseamnă că *nu* există analogii între cercetarea ştiinţifică privind natura aurului şi cercetarea etică sau cea filosofică. În etică, de pildă, pornim de la judecăţile că actele individuale sunt bune sau rele (am putea să le asemănăm rapoartelor de observaţie) şi ajungem să formulăm maxime (nu generalizări excepţionale) bazate pe asemenea judecăţi, însoţite adesea de exemple ilustrative, cum ar fi 'Fiţi buni cu străinii deoarece aţi văzut ce înseamnă să fii străin atunci când aţi fost în Egipt' (o generalizare de nivel scăzut). Asemenea maxime pot modifica judecăţile privind cazurile particulare, generând noi maxime sau modificarea celor vechi. După mii de ani de o asemenea dialectică a maximelor şi judecăţilor privind cazurile individuale, filosofii pot ajunge la alcătuirea unei teorii morale care, la rândul ei, poate modifica atât judecăţile, cât şi maximele.

Aceeaşi procedură este urmată în toate domeniile filosofiei (care este aproape coextensivă cu teoria raţionalităţii). Într-o publicaţie de acum câţiva ani[1] am descris dezideratele pentru un sistem moral, urmându-i pe Grice şi Baker, unde am inclus: (1) cerinţa ca măcar o presupunere fundamentală să fie de larg interes; (2) cerinţa ca sistemul să reziste criticii raţionale; (3) cerinţa ca moralitatea recomandată să poată fi *trăită*.

---

[1] 'Literature, Science and Reflection', *New Literary History*, vol. VII, 1975–6, retipărit în lucrarea mea *Meaning and the Moral Sciences*, Routledge and Kegan Paul, 1978.

Modul în care se poate ajunge la o înţelegere mai bună a naturii raţionalităţii – singurul mod pe care îl ştim – este dezvoltarea unei teorii filosofice mai bune despre raţionalitate. (Un proces fără sfârşit; dar care aşa trebuie să fie). Dezideratele pe care le-am formulat pentru un sistem moral, fără a fi schimbate, pot servi ca deziderate pentru o metodologie sau un sistem de proceduri raţionale în orice domeniu uman major. În filosofia analitică cele mai importante tentative de înţelegere mai bună a naturii raţionalităţii de acest fel vin din partea filosofilor ştiinţei şi, din acest efort, au rezultat două tendinţe importante.

Pozitivismul logic

În ultimii 50 de ani, cea mai clară manifestare a tendinţei de a gândi metodele de 'justificare raţională' ca fiind date printr-o listă sau un canon (deşi câteodată s-a admis că acestea nu pot fi formalizate în întregime) a reprezentat-o mişcarea numită pozitivismul logic. Nu numai lista sau canonul considerau pozitiviştii că ar fi de ajutor 'logicienilor ştiinţei' (cum îi numeau pe filosofi) pentru a descrie exhaustiv 'metoda ştiinţifică'; ci, dacă, în conformitate cu pozitivismul logic, 'metoda ştiinţifică' se suprapune raţionalului şi verificabilitatea este aceea prin care această metodă conferă sens ('Sensul unui enunţ este metoda sa de verificare'), lista sau canonul ar determina şi care sunt judecăţile care au conţinut cognitiv. Judecăţile testabile prin metodele din listă (metodele matematicii, logicii şi ştiinţelor empirice) ar determina sensul; toate celelalte judecăţi, cum menţionează pozitiviştii, sunt 'pseudo-judecăţi' sau non-sensuri deghizate.

Ar fi util să reamintim că criteriul pozitivist logic pentru semnificaţie se *respinge pe sine*, deoarece, nu este nici (a) analitic (termen folosit de pozitivismul logic pentru logică şi matematică), nici (b) empiric testabil. Ciudat că această obiecţie a avut un impact redus asupra pozitiviştilor logici şi nu a fost un impediment în dezvoltarea mişcării. Cred că neglijarea acestei mici neclarităţi filosofice a fost o mare eroare; obiecţia nu este numai corectă, dar conţine şi adânci învăţăminte care se extind dincolo de pozitivismul logic.

Perspectiva pe care doresc să o dezvolt depinde de următoarea observaţie: formele de verificare permise de pozitiviştii logici au fost *instituţionalizate* de societatea modernă. Ceea ce putem 'verifica' în sens pozitivist poate fi verificat ca fiind corect (în sens non-filosofic sau pre-filosofic) sau probabil corect sau având succes, după caz; iar recunoaşterea publică a corectitudinii sau a succesului unei teorii exemplifică, întăreşte şi determină normele de raţionalitate dintr-o cultură.

Pe de o parte, paradigma *originală* a pozitivismului logic privind verificarea nu se suprapune peste modul de recunoaştere publică. În lucrarea lui Carnap *Der Logische Aufbau der Welt*, verificarea, în cele din urmă, este privată ca bazată pe senzaţii a căror calitate subiectivă sau 'conţinut' era 'incomunicabil'. Dar, sub argumentarea lui Neurath, Carnap a acceptat o formă de verificare mai publică sau intersubiectivă.

Popper a subliniat ideea că predicţiile ştiinţifice se confruntă cu 'enunţuri de bază' cum ar fi 'talerul din dreapta al balanţei este mai jos' care sunt acceptate public chiar dacă nu pot fi 'dovedite', spre satisfacţia scepticilor. El a fost criticat pentru că a folosit un limbaj 'convenţionalist', acceptarea unui enunţ de bază apare drept o

convenţie sau o decizie socială; dar după cum cred, convenţionalismul în gândirea lui Popper înseamnă recunoaşterea caracterului *instituţionalizat* al normelor implicite la care apelăm în judecăţile perceptuale obişnuite. Răspunsul la provocarea sceptică de a 'dovedi' asemenea enunţuri precum 'Stau în picioare pe podea' constă în invocarea normelor sociale cerute pentru a cădea de acord cu asemenea judecăţi în circumstanţe date.

Wittgenstein a argumentat că fără asemenea norme publice, norme împărtăşite de un grup care constituie o 'formă de viaţă', limbajul şi gândirea ar fi imposibile. Pentru Wittgenstein, este absurd să întrebăm dacă verificarea instituţionalizată despre care am amintit este realmente 'justificatoare'. În *Despre certitudine* Wittgenstein remarcă faptul că filosofii pot produce sute de justificări epistemologie pentru 'pisicile nu cresc în copaci' – dar *nici una* dintre ele nu pleacă de la ceva care este mai sigur (tocmai în sensul instituţionalizat pentru 'sigur') decât faptul că pisicile nu cresc în copaci.

Scepticii au îndoieli nu numai asupra judecăţilor perceptuale, ci şi asupra inducţiilor obişnuite. Hume, a cărui distincţie între ceea ce este *raţional* şi ce este *rezonabil* nu am luat-o în considerare, ar spune că nu există dovadă *raţională* că va ninge (sau măcar că este probabil să ningă) în Statele Unite în această iarnă (deşi ar adăuga că ar fi foarte nerezonabil să ne îndoim asupra acestui fapt). Răspunsul pentru un sceptic care s-ar îndoi că va ninge în Statele Unite în această iarnă, face apel la norme sociale potrivit cărora suntem de acord cu asemenea 'inducţii', la fel cum cădem de acord că cineva stă în picioare pe podea şi că braţele unei balanţe sunt echilibrate.

Când ajungem la o teorie de un nivel înalt în ştiinţele exacte, oamenii reacţionează oarecum diferit. Omul de rând nu poate 'verifica' teoria specială a relativităţii. Omul obişnuit, chiar în zilele noastre, *învaţă* o asemenea teorie, sau matematica de care are nevoie pentru a o înţelege, chiar dacă nu este novice în anumite probleme de fizică. Omul obişnuit lasă pe seama omului de ştiinţă cercetarea asupra unei teorii de acest tip. Datorită instabilităţii teoriilor ştiinţifice, un om de ştiinţă nu este în măsură să considere o teorie, chiar dacă are succes, cum este teoria relativităţii speciale, ca adevărată *tout court*. Judecata comunităţii ştiinţifice este că teoria relativităţii speciale are succes – de fapt ca şi electrodinamica cuantică, un succes fără precedent – deoarece produce 'predicţii de succes' şi deoarece este 'susţinută de un mare număr de experimente'. Aceste judecăţi sunt, de fapt, oferite celorlalţi membri ai societăţii. Diferenţa dintre acest caz şi cel al normelor instituţionalizate de verificare (separat de domeniul adjectivului 'adevărat') este dată de rolul special pe care îl au experţii; dar aici nu este altceva decât un exemplu de diviziune a muncii intelectuale (şi nu pentru a susţine relaţii de autoritate intelectuală) în societate. Judecata că teoria relativităţii şi electrodinamica cuantică sunt 'cele mai de succes teorii fizice pe care le avem' este făcută de autorităţile pe care le-a propulsat societatea şi a căror autoritate este recunoscută prin practici sau ceremonii şi astfel, instituţionalizată.

Recent, mi-am dat seama că Wittgenstein ar putea avea în vedere că numai *judecăţile* care ar putea fi verificate într-un mod instituţionalizat pot să fie cât de cât adevărate

(sau corecte, ori justificate). Nu vreau să sugerez că orice filosof a susţinut punctul de vedere că toate lucrurile care contează în societatea noastră ca 'justificate' sunt astfel. Filosofii disting între instituţii care sunt constitutive pentru conceptele noastre şi cele care au alt statut, deşi este dificil de spus cum se face o asemenea distincţie. Sugestia mea este că Wittgenstein credea că există o submulţime de norme de verificare instituţionalizate care determină ce este corect să se spună în diferitele 'jocuri de limbaj' şi ce este greşit şi că nu există corectitudine şi eroare obiectivă dincolo de acestea. Deşi o asemenea interpretare trece dincolo de ceea ce spune Wittgenstein – de pildă, accentuează necesitatea acordului în judecăţile noastre pentru a avea concepte – totuşi nu sunt sigur că are dreptate. Este prea vag cine suntem 'noi' când Wittgenstein vorbeşte de 'judecăţile' noastre; în plus, nu ştiu dacă 'formele de viaţă' pe care le invocă sunt corespunzătoare normelor instituţionalizate. Dar această interpretare mi-a fost sugerată citind *Prelegerile şi conversaţiile* lui Wittgenstein. Aici Wittgenstein respinge atât psihanaliza, cât şi teoria evoluţiei a lui Darwin (deşi spre deosebire de pozitivişti el nu priveşte asemenea limbaje ca lipsite de sens şi are admiraţie pentru 'subtilitatea' lui Freud). Perspectiva lui Wittgenstein asupra psihanalizei (pe care o consideră un 'mit') nu înseamnă mult, odată ce numeroşi oameni susţin – în mod eronat în opinia mea – că psihanaliza nu are sens. Dar respingerea teoriei evoluţiei este cu totul neaşteptată[2].

[2] În ce priveşte evoluţia, Wittgenstein afirma: 'Temeiul este destul de nesigur. Nu se putea adopta atitudinea "Nu ştiu. Este o ipoteză interesantă care ar putea fi eventual confirmată"', *Lectures on Aesthetics*, p. 26, în Cyril Burret (ed.) *L. Wittgenstein: Lectures and Conversations*,

Wittgenstein compară teoria lui Darwin cu teoriile fizicii ('Unul dintre cele mai importante aspecte ale explicaţiei este că ar trebui să funcţioneze aşa încât să ne permită predicţii. Fizica este conectată cu ingineria. Podul nu va cădea'. (*Prelegeri despre estetică*, p. 25)). Pe de altă parte, oamenii sunt convinşi 'pornind de la fundamente extrem de slabe'. 'La final uitaţi toate problemele legate de verificare, fiţi siguri că aşa se petrec lucrurile'.

Din nou, discuţiile aprinse despre analiticitate purtate în anii 1950 mi se par a fi legate de dorinţa filosofilor de a găsi fundamente obiective, dincolo de controverse pentru argumentele lor. 'Analiticitatea', adică doctrina adevărului în virtutea sensului, ajunge să fie respinsă pentru că s-a abuzat de ea. Dar de ce au fost tentaţi filosofii să anunţe că atât de multe lucruri care sunt 'reguli de limbaj' într-un mod neinteligibil, sau consecinţe ale acestor reguli, ar fi analitice sau 'necesare conceptual'? Cred că răspunsul îl găsim în ideea că existenţa unei mulţimi definite de *reguli*

Berkeley, University of California Press, 1967. Ce ar însemna ca evoluţia să fie 'bine confirmată', Wittgenstein nu spune, dar paragraful sugerează că se gândeşte la *vederea* apariţiei speciilor ('Oare cineva a văzut un asemenea proces? Nu. A văzut cineva întâmplându-se acum? Nu. Evidenţa dată de existenţa hibrizilor este numai o picătură într-un ocean.') Aici este instructiv contrastul dintre atitudinea lui Wittgenstein şi cea a lui Monod: 'teoria evoluţiei prin selecţie, aşa cum a susţinut-o Darwin, cere descoperirea geneticii mendeliene, care a şi avut loc. Acesta este un exemplu, chiar cel mai important, despre ce înseamnă conţinutul unei teorii, conţinutul unei idei... O teorie bună sau o idee bună devine în timp mai extinsă şi bogată decât putea prevedea inventatorul ei. Teoriile pot fi judecate exact după modul în care evoluează, când cele mai multe se dovedesc eronate, deşi nu era de aşteptat să se întâmple astfel' (J. Monod, 'On the Molecular Theory of Evolution', Harre R. (ed.), *Problems of Scientific Revolution: Progress and Obstacles to Progress in Science*, Oxford, 1975).

ale limbajului prin care se poate stabili ce este şi ce nu este raţional are două avantaje, după cum credeau filosofii respectivi: (1) 'regulile limbajului' sunt practici instituţionale constitutive (sau norme care implică asemenea practici) şi au statutul public pe care l-am amintit; (2) în acelaşi timp, se afirma că numai filosofii (nu şi lingviştii) ar putea descoperi aceste lucruri misterioase. Era o idee drăguţă, dar care nu a ţinut mult.

Voi numi *criterială*, orice concepţie despre raţionalitate conform căreia există norme instituţionalizate care definesc ce este şi ce nu este raţional acceptabil. Pozitiviştii logici, Wittgenstein, cel puţin pe baza interpretării probabile pe care am propus-o, cât şi unii dintre filosofii 'limbajului comun'[3] de la Oxford

[3] S-ar putea dezvolta o filosofie a 'limbajului comun' care să nu fie tributară verificării publice şi 'criteriale' a tezelor filosofice dacă s-ar putea dezvolta şi susţine o concepţie în care normele care guvernează practicile lingvistice nu sunt descoperite prin investigaţie empirică. În lucrarea *Must We Mean What We Say*, Stanley Cavell face un pas semnificativ în această direcţie argumentând că asemenea norme pot fi cunoscute printr-o specie de 'introspecţie' pe care o compară cu psihoterapia sau cu cunoaşterea transcendentală propovăduită de fenomenologie. Deşi sunt de acord cu Cavell că o asemenea cunoaştere de sine în calitate de vorbitor nativ care uneori se abate de la norme nu este o cunoaştere externă inductivă – nu pot să ştiu fără a face apel la evidenţă că în dialectul meu de engleză se spune 'mice' şi nu 'mouses' – sunt înclinat să cred că un asemenea mecanism nu conduce la *generalizări* despre corectitudine sau incorectitudine. Dacă afirm (cum face Cavell) că *X* ar trebui să fie ambiguu este o parte a regulii pentru utilizarea corectă a locuţiunilor de forma *X este opţional*, atunci propun o teorie pentru a explica intuiţiile mele despre cazuri specifice fără a da seama de aceste intuiţii. Într-adevăr, ceva asemănător are loc în psihoterapie; dar nu exclud posibilitatea ca autocunoaşterea să poată fi supusă criticii de către ceilalţi, inclusiv criticii care depinde de ipoteze rivale. Dacă un asemenea criticism este permis, atunci activitatea prin care sunt descoperite asemenea norme devine asemănătoare celei din

împărtăşesc o concepţie criterială asupra raţionalităţii chiar dacă diferă între ei în alte privinţe, cum ar fi considerarea judecăţilor neverificabile ca fără sens şi dacă judecăţile etice sunt 'necesare conceptual' sau nu.

Dificultatea la care m-am referit la început, care afectează principiul pozitivist logic al verificabilităţii, este *adâncă* deoarece respinge orice tentativă de argumentare pentru concepţia criterială asupra raţionalităţii, respectiv, pentru că respinge teza că nimic nu este raţional verificabil dacă nu este criterial verificabil.

Deşi filosofii amintiţi au vorbit adesea de parcă argumentele lor au acelaşi tip de finalitate precum demonstraţiile matematice sau experimentele fizice, deşi pozitiviştii logici numeau activitatea lor *logică* a ştiinţei, deşi adepţii lui Wittgenstein manifestă o aroganţă incredibilă faţă de cei care nu 'văd' că întreaga filosofie a celor dinainte de Wittgenstein sau a adversarilor lui Wittgenstein ar fi lipsită de sens, deşi filosofii limbajului comun privesc argumentele celorlalţi, ca şi pe filosofii limbajului necomun, ca pe nişte plângăcioşi (de parcă erorile filosofice ar fi greşeli la un test de aritmetică), nici o poziţie filosofică de oarecare importanţă nu poate fi verificată în mod concluziv şi definitiv. Pe scurt, dacă este adevărat că numai judecăţile pe care le putem verifica criterial pot fi raţional acceptabile, tocmai o asemenea judecată nu poate fi verificată criterial aşa încât nu este raţional acceptabilă. Dacă există un asemenea lucru precum raţionalitatea – şi ne angajăm să credem într-o noţiune de raţionalitate prin

---

istorie sau ştiinţele sociale – domenii în care, aşa cum am argumentat, 'metoda ştiinţifică tradiţională' are puţine de spus. (Vezi lucrarea mea *Meaning and the Social Sciences*, Routledge and Kegan Paul, 1978). În orice caz, oricare i-ar fi statutul, nu văd nici un motiv pentru a crede că normele de utilizare a *limbii* sunt decisive în ce priveşte 'acceptabilitatea raţională', 'justificarea', 'bine confirmarea' etc.

activităţi de vorbire şi argumentare – atunci teza că normele instituţionale ale culturii determină instanţe ale ei se respinge singură. Deoarece nici un asemenea argument nu poate fi sigur corect sau măcar probabil corect, numai pe baza acestor norme.

Nu cred că argumentarea şi justificarea raţională sunt imposibile în filosofie, ci mai degrabă că am fost nevoiţi să recunoaştem ceva care este probabil evident pentru un om obişnuit dacă nu pentru un filosof, respectiv că nu putem apela la normele *publice* pentru a decide ce este şi ce nu este raţional argumentat şi justificat în filosofie. Afirmaţia făcută adesea că filosofia este o 'analiză conceptuală' prin care conceptele însele determină corectitudinea unui argument ajunge, dacă o combinăm cu doctrina potrivit căreia conceptele sunt norme sau reguli ale practicilor lingvistice publice, doar acoperire pentru afirmaţia că orice justificare raţională în filosofie este criterială şi că adevărul filosofic este dincolo de eroare aidoma adevărului ştiinţific *demonstrabil public*. Un asemenea punct de vedere mi se pare nerezonabil în lumina întregii istorii a acestui subiect, incluzând istoria recentă.

Ce este valabil pentru argumentarea filosofică este valabil şi pentru cea religioasă sau ideologică. O dispută între un liberal inteligent şi un marxist inteligent va avea acelaşi caracter cu o dispută filosofică, chiar dacă faptele empirice joacă un rol mai important. În religie, politică sau filosofie întâlnim aceeaşi situaţie, peste tot se argumentează şi se critică argumentele celorlalţi. Chiar în ştiinţele care nu sunt exacte, în istorie, în sociologie şi în psihologia clinică, întâlnim argumentare. Pozitiviştii logici au limitat descrierea lor la

'metoda ştiinţifică' dar o asemenea limitare nu *exclude* o aplicabilitate mai largă (vezi Capitolul 8).

Pozitiviştii, aşa cum am amintit, *admit* că principiul verificabilităţii este 'lipsit de sens cognitiv'. Ei susţin că avem de-a face cu o presupunere, astfel că nu este nici adevărat, nici fals. Pe de altă au argumentat pentru această presupunere, iar argumentele nu erau (şi nu puteau să fie) decisive[4]. Astfel că obiecţia rămâne.

În rezumat, pozitivismul logic, alături de Wittgenstein (şi poate Quine), a *elaborarat o filosofie care nu lasă loc de activitate raţională pentru filosofie.* De aceea, aceste opţiuni se resping singure; de asemenea, iată de ce mica dificultate la care m-am referit anterior reprezintă un

---

[4] Cel mai slab argument dezvoltat pentru apărarea principiului verificării construit ca o presupunere era că în acest fel se 'explica' noţiunea 'preanalitică' de sens. (Pentru amănunte asupra acestei chestiuni vezi lucrarea mea 'How Not to Talk about Meaning', în *Mind, Language and Reality, Philosophical Papers*, vol. 2, Cambridge U.P., 1975). Reichenbach apără o formă a principiului verificării (în *Experience and Prediction*) deoarece *păstrează diferenţele de sens relevante în ce priveşte comportamentul.* Împotriva unei obiecţii imediate (că convingerea non-empirică în divinitate – Reichenbach are în vedere credinţa egiptenilor în pisici – poate influenţa comportamentul) Reichenbach răspunde propunând ca expresia 'Pisicile sunt animale divine' să fie interpretată prin 'Pisicile inspiră sentimente de teamă credincioşilor'. Dar o asemenea interpretare *nu* lasă neschimbat comportamentul unui adorator! Cea mai interesantă perspectivă aparţine lui Carnap. Potrivit acestuia, *toate* reconstrucţiile raţionale sunt propuneri. Numai aspectele factuale privesc consecinţele logice şi empirice ale acceptării cutării reconstrucţii raţionale. (Carnap compară 'alegerea' unei reconstrucţii raţionale cu alegerea unui motor pentru un avion). Concluzia pe care a tras-o era că în filosofie se impune toleranţa faţă de reconstrucţiile raţionale divergente. Totuşi, acest principiu al toleranţei, cum îl numeşte Carnap, *presupune* principiul verificării. Pentru a susţine doctrina că nici o reconstrucţie raţională nu este singura *corectă* sau singura care corespunde modului în care sunt lucrurile 'în realitate', este nevoie de doctrina că toate 'aspectele externe' nu au sens cognitiv, care *este* tocmai principiul verificării. Aplicarea principiului toleranţei principiului verificării este circulară.

amănunt semnificativ pentru argumentul numit de unii filosofi 'transcendental': argumentarea cu privire la natura raţionalităţii (sarcina filosofiei *par excellence*) este o activitate care presupune o noţiune de justificare raţională mai largă decât cea pozitivistă, mai largă decât raţionalitatea criterială instituţionalizată.

Anarhism şi autorespingere

Să abordăm acum o perspectivă cu totul diferită. Lucrarea lui Thomas Kuhn *Structura Revoluţiilor Ştiinţifice* (*SRS*) a fermecat un mare număr de cititori şi a înspăimântat pe cei mai mulţi filosofi ai ştiinţei deoarece pune în evidenţă o determinare *iraţională* a acceptării teoriilor ştiinţifice prin folosirea unor termeni precum 'conversie' sau 'răsturnare a Gestalt-ului'. Kuhn face un număr important de observaţii privind teoriile ştiinţifice şi activitatea ştiinţifică. Am pus în evidenţă importanţa unor noţiuni precum *paradigmă*, *ştiinţă normală* şi *revoluţie ştiinţifică* în alte ocazii; aici doresc să mă opresc la aspecte din cartea lui Kuhn faţă de care nu sunt de acord, pe care le-am descris drept 'relativismul extrem a lui Kuhn'.

Interpretarea care a atras cititorii cei mai documentaţi este că nu există nici un fel de justificare *raţională* în ştiinţă, ci *numai* schimbări şi conversiuni ale Gestalt-ului. Kuhn a respins această interpretare a *SRS* şi a introdus noţiunea de 'raţionalitate neparadigmatică' prin care se apropie de ceea ce eu numesc 'raţionalitate necriterială'.

Tendinţa celor mai mulţi cititori a fost să creadă că unele aspecte din *SRS* sunt reluate în lucrarea lui Paul Feyerabend *Împotriva metodei*. Feyerabend, la fel cu Kuhn,

accentuează maniera în care diferite culturi şi epoci produc paradigme diferite ale raţionalităţii. El consideră că elementele pe care le numim *iraţionale* sunt determinante pentru concepţia *noastră* asupra raţionalităţii ştiinţifice. Ca urmare, deşi nu o spune direct, Feyerabend sugerează că perspectiva modernă ştiinţifico-tehnologică despre raţionalitate este respinsă prin propriile sale standarde. (Cred că găsesc o tentă similară şi la M. Foucault). Totuşi, el merge mai departe decât Kuhn sau Foucault în sugestia că inclusiv superioritatea instrumentală a ştiinţei noastre ar putea fi o glumă. Vracii pot face mai mult pentru a alina durerea decât doctorii, afirmă Feyerabend.

Nu asupra acestor afirmaţii prea radicale vreau să mă opresc, deşi ele sunt motivul pentru care Feyerabend îşi numeşte poziţia 'anarhism'. Aş dori să abordez o afirmaţie a lui Kuhn apărută atât în *SRS*, cât şi în alte lucrări şi pe care Feyerabend o reia în *Împotriva metodei* şi alte scrieri mai tehnice. Este vorba de teza *incomensurabilităţii*. Această teză, la fel cu aceea despre sens şi verificare a pozitiviştilor logici, este una care se autorespinge. Pe scurt, *două* dintre cele mai influente filosofii ale ştiinţei din secolul al XX-lea, cu siguranţă cele care au trezit cel mai mare interes chiar printre oamenii de ştiinţă, singurele despre care cititorii educaţi au auzit, se anulează singure. De bună seamă, ca filosof al ştiinţei, găsesc tulburător un asemenea rezultat. Vom expune pe scurt situaţia.

Teza incomensurabilităţii susţine că termenii folosiţi într-o altă cultură, cum ar fi termenul 'temperatură' utilizat de un om de ştiinţă din secolul al XVII-lea, nu pot fi comparaţi în ce priveşte sensul sau referinţa cu termenii sau expresiile *noastre*. După cum susţine Kuhn, oamenii de

ştiinţă adepţi ai unor paradigme diferite locuiesc în lumi diferite. 'Electron' cum era folosit prin 1 900 se referea la obiecte dintr-o 'lume'; astăzi se referă la obiecte din altă 'lume'. Teza se aplică atât limbajului observaţional, cât şi celui teoretic; într-adevăr, după Feyerabend, limbajul comun este pur şi simplu o teorie falsă.

Dacă teza ar fi adevărată, ar urma că traducerea dintr-o limbă în alta sau între stadii diferite ale aceleiaşi limbi ar fi imposibilă. Odată ce nu putem interpreta 'zgomotele' organismelor în nici un fel, nu avem nici un temei să le considerăm drept *gânditori*, *vorbitori* sau chiar *persoane*. Pe scurt, dacă Feyerabend (sau Kuhn pe când susţinea incomensurabilitatea) au dreptate, membrii altor culturi, inclusiv oamenii de ştiinţă din secolul al XVII-lea, ar fi văzuţi de către noi numai ca nişte animale care răspund la stimuli (inclusiv prin zgomote care se aseamănă în mod curios cu engleza sau italiana). A spune că Galileo avea noţiuni incomensurabile şi totodată a le descrie în detaliu este total incoerent. Această problemă este pusă într-un eseu despre Feyerabend aparţinând lui Smart[5]:

> Desigur că este un fapt neutru că, pentru a-l vedea pe Mercur avem nevoie de un punct de observaţie deasupra acelui arbore, să spunem, şi nu, aşa cum cere teoria lui Newton, deasupra acelui horn. Se poate vorbi de arbori, hornuri şi telescoape independent de alegerea teoriei lui Newton sau a lui Einstein. Cu toate acestea, Feyerabend ar fi de acord că folosim optica potrivit geometriei lui Euclid şi nu celei relativiste pentru

[5] J. J. C. Smart, 'Conflicting Views about Explanation', în R. Cohen şi M. Wartofski (eds.), *Boston Studies in the Philosophy of Science, Volume II: in Honour of Philipp Franck* (New York, Humanities Press, Inc., 1965).

calculul legat de telescop. El ar spune că nu avem de-a face cu un adevăr real cu privire la relaţia dintre telescop, arbore şi horn, dar discuţia privind testarea observaţională a relativităţii generale este legitimă dacă ştim că temeiurile teoretice ale predicţiilor noastre nu sunt afectate (dincolo de limitele erorilor de observaţie) şi serveşte calculelor.

Dar problema cu tentativa de salvare a lui Smart este că eu trebuie să-i înţeleg pe *unii* dintre euclidienii nerelativişti pentru a vedea dacă fac aceeaşi predicţie. Dacă *orice cuvânt are o semnificaţie diferită* în ce sens poate fi 'neafectată' o predicţie? Cum am putea traduce chiar particulele logice (cuvinte precum 'dacă-atunci', 'nu' etc.) în italiana secolului al XVII-lea dacă nu putem găsi un manual care să permită legătura dintre limba italiană din secolul al XVII-lea cu româna modernă, care să dea un fel de înţeles sistemic corpusului din secolul al XVII-lea atât în sine, cât şi în aplicaţiile extra-lingvistice? Chiar dacă aş fi un vorbitor care foloseşte ambele teorii (cum are în vedere Smart) cum pot justifica legătura pe care o fac între cuvintele teoriei lui Newton şi cele ale teoriei relativităţii?

Acum este momentul să discutăm asupra observaţiilor privind traducerea şi sensul făcute de Quine şi Davidson. Dacă se acceptă că putem găsi o schemă de traducere care 'funcţionează' în cazul unui text din secolul al XVII-lea, cel puţin în contextul fixat de interesele noastre şi prin utilizarea căruia traducerea poate fi dusă la bun sfârşit, ce înţeles trebuie să aibă *în acel context* că traducerea nu 'captează' sensul originar? Nu înseamnă că am avea criterii pentru sinonimie deosebite de schema traducerii şi cerinţe explicite sau implicite pentru adecvarea empirică. Afirmaţia că traducerea nu reuşeşte să capteze sensul sau referinţa exactă a originalului poate fi înţeleasă dacă se poate face o traducere

mai bună; dar este iluzoriu să spui că orice schemă de traducere posibilă eşuează în surprinderea înţelesului sau referinţei reale. Sinonimia există numai ca o relaţie, sau, mai curând, ca o familie de relaţii, fiecare destul de vagi, care sunt folosite pentru a echivala diferite expresii în scopul interpretării lor. Teza că ar exista vreun fel de 'sinonimie reală' pe lângă cea rezultată din practica interpretării trebuie demascată drept un mit.

Să presupunem că cineva spune despre cuvântul german 'Rad' că poate fi tradus prin 'roată'. Dacă cineva afirmă că traducerea nu este perfectă, aşteptăm să arate cum poate fi îmbunătăţită, suplimentată printr-o explicaţie sau în alt mod. Dar dacă acesta afirmă că 'Rad' poate fi tradus prin 'roată' dar de fapt nu se referă la roţi sau la alt obiect ce poate fi recunoscut prin sistemul conceptual propriu, ce concluzie tragem? A spune despre cuvântul *A* că poate fi tradus prin 'roată', sau altfel, înseamnă a spune că, extinzând relaţia presupusă prin traducere, *A se referă* la roţi.

Poate că motivul pentru care teza incomensurabilităţii intrigă aşa de mult, dincolo de incoerenţa la care conduce, este tendinţa de a confunda conceptul şi concepţia. Distincţia analitic/sintetic este prea vagă pentru a putea aprecia cât este de vagă; dar orice interpretare implică o asemenea distincţie chiar dacă este relativă la interpretarea însăşi. Atunci când traducem un cuvânt precum *temperatură* echivalăm referinţa, ca şi înţelesul expresiei traduse, cu cel al propriului termen de 'temperatură', cel puţin aşa cum este folosit în acel context. (De bună seamă că există instrumente diferite pe care le putem utiliza, cum ar fi dicţionarele, pentru a delimita modul în care 'temperatură' este utilizat, sau orice alt cuvânt, într-un context.) În acest mod, echivalăm conceptul în chestiune cu propriul nostru 'concept' de

temperatură. Dar o asemenea operaţie este compatibilă cu faptul că oamenii de ştiinţă din secolul al XVII-lea pot avea o *concepţie* diferită despre temperatură, adică o mulţime diferită de convingeri cu privire la aceasta şi natura sa decât noi, o 'imagine cognitivă' diferită, ca şi convingeri diferite despre multe alte lucruri. Dar toate acestea nu fac imposibilă traducerea cum s-a presupus uneori; dimpotrivă, nu se poate spune că există diferenţe de concepţii şi care este această diferenţă dacă traducerea nu este posibilă.

Dar s-ar putea ridica întrebarea, cum ştim că o traducere funcţionează odată ce concepţiile sunt întotdeauna diferite? Răspunsul la această problemă, aşa cum a fost dat de numeroşi gânditori de la Vico până astăzi, este că succesul interpretativ nu cere coincidenţa convingerilor celui tradus cu ale noastre, ci cere ca ele să ne fie *inteligibile*. Aceasta este baza pentru toate maximele deschiderii interpretative sau 'beneficiul îndoielii', cum ar fi: 'interpretează astfel ca ei să devină credincioşi ai adevărului şi iubitori ai binelui' sau 'interpretează astfel încât convingerile să ajungă rezonabile potrivit cu ideile şi experienţele lor' sau directiva lui Vico de a maximiza *umanitatea* persoanei interpretate. Este un fapt al experienţei umane că într-o lume unde interacţionează culturi diferite, unde individul trebuie să se adapteze schimbărilor, acesta trebuie să fie în măsură să interpreteze rostirile, convingerile şi dorinţele celuilalt, să le confere un înţeles.

Kuhn şi Feyerabend nu surprind atunci când resping orice *convergenţă* în cunoaşterea ştiinţifică. O dată ce nu vorbim despre aceleaşi lucruri ca şi predecesorii noştri, nu putem adăuga mai multă cunoaştere cu privire la aceleaşi obiecte microscopice sau macroscopice. Kuhn argumentează că ştiinţa progresează numai sub aspect instrumental; suntem în măsură să transportăm mai bine

oamenii dintr-o parte în alta etc. Dar o asemenea perspectivă este şi ea incoerentă. Chiar numai asemenea locuţiuni precum 'transportul oamenilor dintr-un loc în altul' cer un grad de fixare a referinţei, altfel cum am putea înţelege noţiunea de succes instrumental?

Argumentul pe care tocmai l-am invocat este strâns legat de celebrele argumente privind precondiţiile pentru cunoaşterea empirică formulate de Kant. Replicând obiecţiei că viitorul ar putea fi complet haotic, deoarece orice inducţie pe care am invoca-o ar fi eronată, Kant subliniază că dacă există vreun viitor – vreun viitor *pentru noi*, adică orice viitor ne-am putea imagina pentru a spune dacă predicţiile noastre sunt adevărate sau false – atunci, de fapt, există numeroase regularităţi care nu pot fi violate. Altfel, de ce am vorbi despre un *viitor*? De pildă, când ne imaginăm bile care ies dintr-o urnă într-o ordine neregulată, uităm că *nu am putea afirma că avem de-a face cu bile* sau *în ce ordine apar*, fără a depinde de numeroase regularităţi. *Compararea* presupune că există comensurabilitate.

Există un răspuns posibil din partea lui Kuhn sau Feyerabend la astfel de critici, dar nu este unul care să-i facă fericiţi, şi care ar putea fi introdus printr-o dihotomie între observaţional şi teoretic. Ei ar putea accepta comensurabilitatea şi chiar convergenţa cu privire la faptele observaţionale şi să restrângă incomensurabiltiatea la vocabularul teoretic. Chiar în acest caz ar apărea probleme (de ce să nu exprimăm sensurile termenilor teoretici prin termeni observaţionali folosind o metodă asemănătoare celei dezvoltate de Ramsey?). Dar Kuhn şi Feyerabend resping o asemenea variantă pe motivul că avem nevoie de principii ale deschiderii interpretative, atât în 'limbajul observaţional', cât şi în 'limbajul teoretic'. Să ne oprim, de

pildă, la cuvântul 'iarbă'. Vorbitori diferiţi, în funcţie de locul şi momentul în care se află, au prototipuri diferite de iarbă (iarba are diferite culori şi forme diferite în diverse locuri) şi concepţii diferite despre iarbă. Chiar dacă toţi vorbitorii ar fi de acord că iarba este alcătuită din plante, ar putea exista excepţii, deoarece concepţia despre *plantă* se modifică, astăzi presupune fotosinteza, spre deosebire de acum două sute de ani. Fără deschiderea interpretativă care ne permite să echivalăm 'plantă' de acum 200 de ani cu 'plantă' astăzi (cel puţin pentru contextele obişnuite) şi 'iarbă' de acum 200 de ani cu 'iarbă' de astăzi, nu s-ar putea face nici o judecată cu privire la referinţa acestui cuvânt acum 200 de ani. Nu numai cuvintele pentru tipuri naturale sunt dependente de interpretarea după principiul deschiderii interpretative; cuvântul pentru un artefact, cum este 'pâine', pune exact aceleaşi probleme. În plus, fără deschidere interpretativă nu putem echivala un simplu termen pentru culoare, cum este 'roşu' la diferiţi vorbitori. Discursul trebuie interpretat ca un întreg; iar interpretarea termenilor de observaţie este dependentă de interpretarea termenilor 'teoretici', aşa cum interpretarea acestora din urmă depinde de cea a primilor.

Din nou am ajuns la argumentul transcendental. Ne angajăm prin concepţiile noastre fundamentale să tratăm nu numai timpul nostru prezent, ci şi trecutul nostru, strămoşii noştri şi membrii altor culturi trecute şi prezente, ca *persoane*; asta înseamnă, cum am arătat, să le atribuim accesul la aceleaşi referinţe şi la aceleaşi concepte, cu toate că trebuie să le atribuim şi *concepţii* diferite. Nu împărtăşim numai conceptele şi obiectele cu ceilalţi când apreciem că interpretările au succes, ci de asemenea, concepţiile despre ce este rezonabil, despre ce este natural etc., deoarece întreaga justificare pentru schema inter-

pretativă, vă reamintesc, este că aceasta traduce comportamentul celorlalţi într-un mod cât mai rezonabil. În plus, pe lângă imaginile cu privire la cunoaştere şi concepţii despre raţionalitate, împărtăşim cu toţii, cu toate diferenţele culturale, un bogat fond de presupuneri şi convingeri despre ceea ce este rezonabil.

De ce relativismul este inconsistent

Faptul că relativismul (total) este inconsistent este un truism printre filosofi pentru că a *susţine* un punct de vedere, simultan cu teza că nici un punct de vedere nu este mai justificat sau mai bun decât oricare altul, este contradictoriu *în mod evident*. Alan Garfinkel a pus această problemă într-un mod spiritual. În discuţiile cu studenţii din California a spus odată, răspunzând solicitărilor lor: 'Nu putem avea aceeaşi formaţie, dar ştiu că relativismul *nu este adevărat pentru mine*...' Dacă orice punct de vedere este la fel de bun cu oricare altul, atunci de ce *punctul de vedere că relativismul este fals nu este la fel de bun cu celelalte*?

Abundenţa de doctrine relativiste promovate astăzi (de către gânditori foarte inteligenţi) arată că această respingere simplă nu este suficientă. De ce ar trebui un relativist inteligent să fie de acord că orice punct de vedere este tot atât de adevărat (pentru el) cu al său? El nu poate să vă prevină (sau Alan Garfinkel) spunând că punctul său de vedere *nu este adevărat pentru voi* (sau justificat pentru voi etc.): dar el are discernământ despre el, el poate riposta că adevărul pentru tine este cu mult mai puţin proeminent (*pentru el*) decât adevărul pentru el. Care concept, privind ceva mai proeminent decât el însuşi, există de fapt? În acest

caz, de ce este inconsistent să tratăm *adevărul*, *justificarea* etc. drept noţiuni *relative*?

Răspunsul ar fi: este inconsistent, dar avem nevoie de un argument mai elaborat decât cel (de bună seamă, interesant) produs de Garfinkel. Trebuie evidenţiat că, dacă totul este relativ, atunci şi relativul este relativ. Dar pentru asta, se cer adăugate unele explicaţii.

Platon a fost, poate, primul filosof care a utilizat un asemenea argument împotriva lui Protagoras. Protagoras (un relativist fanatic) afirma că, dacă eu spun *X*, ar trebui să spun 'Eu *cred X*'. Astfel, când spun 'Zăpada este albă', după Protagoras, eu afirm de fapt că Hilary Putnam crede că zăpada este albă, iar ceea ce afirmă Robert Nozick prin aceeaşi rostire este că Robert Nozick crede că zăpada este albă. O prezentare mai sofisticată a aceleiaşi idei ar fi că, spunând 'Zăpada este albă', folosesc această rostire pentru a spune că *zăpada este albă* este adevărată pentru *mine*, în vreme ce, atunci când Robert Nozick spune aceleaşi cuvinte, el ar afirma de fapt că *zăpada este albă* este adevărată pentru *el* (sau cel puţin ar lua în considerare faptul că judecata ar fi corectă numai în cazul în care ar fi adevărată pentru el). Urmează (din perspectiva lui Protagoras) că nici o rostire nu are acelaşi sens pentru mine şi pentru oricine altcineva; există, cum am văzut anterior, o conexiune strânsă între relativism şi incomensurabilitate. Contraargumentul lui Platon era că, dacă orice judecată *X* înseamnă 'Eu cred că *X*', atunci ar trebui (din perspectiva lui Protagoras) să spunem:

(1) Eu cred că eu cred că zăpada este albă.

Dar adăugarea lui 'eu cred' poate fi continuată! Astfel, din perspectiva lui Protagoras, înţelesul pentru 'Zăpada este albă' nu este (1) ci

(2) Eu cred că eu cred că eu cred... (cu infinit de mulţi 'Eu cred') că zăpada este albă.

Acest argument platonician este o *reductio ad absurdum*. Totuşi, nu este atât de bun pe cât pare. De ce ar trebui ca Protagoras să fie de acord că această analiză i se aplică? Nu este obligatoriu că analiza *trebuie* să se aplice de un număr finit de ori, ci numai că *poate* să se aplice de un număr infinit de ori. Dar Platon a observat ceva mai profund.

Când ne întâlnim prima dată cu relativismul, ideea pare destul de simplă. În forma lui naturală, relativismul susţine că orice persoană (sau într-o formulare sociologică modernă, orice cultură, iar uneori orice discurs) are propria perspectivă, propriile standarde şi presupoziţii şi că adevărul (cât şi justificarea) sunt relative la *acestea*. Unii ar descrie o asemenea concepţie spunând că '*X* este adevărat relativ la acestea' este un adevăr absolut.

Structuraliştii moderni, cum este Foucault, scriu de parcă justificarea *relativă la discurs* este ea însăşi absolută – nicidecum relativă. Dar dacă judecata de forma '*X* este adevărată (justificat) relativ la persoana *P*' este ea însăşi adevărată sau falsă *absolut*, atunci există, în cele din urmă, o noţiune absolută de adevăr (sau de justificare) nu numai 'adevărat pentru mine' sau 'adevărat pentru Nozick', 'adevărat pentru tine' etc. Un relativist *total* ar trebui să spună că dacă (sau nu) *X* este relativ la *P* este relativ în sine. La acest moment surprindem cum relativismul începe să se clatine, aşa cum a observat Platon.

Strategia lui Platon de respingere a relativismului nu a fost urmată decât recent, fiind extinsă cu strălucire de Wittgenstein cu ocazia argumentului limbajului privat (amintit în Capitolul 3).

Cei mai mulţi comentatori înţeleg argumentul limbajului privat ca îndreptat împotriva teoriei adevărului copie. Demonstraţia lui Wittgenstein că teoria similarităţii privind referinţa nu funcţionează nici pentru referirea la senzaţii este cu siguranţă o parte din atacul său împotriva realismului metafizic. Eu prefer să interpretez argumentul ca o *pereche* de argumente tradiţionale (cel puţin Kant le-ar aproba pe ambele) împotriva a două poziţii, cea realistă şi cea relativistă pentru că încercarea de a interpreta întregul argument ca antirealist îl face să pară mai degrabă ticluit.

Forma de relativism pe care o atacă Wittgenstein este cunoscută drept 'solipsism metodologic'. Un 'solipsist metodologic' este un non-realist sau un 'verificaţionist' care acceptă că adevărul trebuie înţeles într-un mod legat de acceptabilitatea raţională, dar care susţine că orice justificare trebuie redusă la termenii experienţei pe care fiecare dintre noi o cunoaşte în mod privat. Cu alte cuvinte eu am cunoaşterea *proprie* a experienţei *mele* şi prin aceasta verific dacă zăpada este albă, iar Bob Nozick are cunoaşterea *sa* a experienţei *sale* şi prin aceasta el verifică dacă zăpada este albă: orice judecată are un înţeles diferit pentru fiecare gânditor.

Consider că argumentul lui Wittgenstein este unul excelent împotriva relativismului în general. Argumentul constă în aceea că relativistul nu poate, în cele din urmă, să confere înţeles distincţiei între *a avea dreptate* şi *a crede că ai dreptate*; de aici rezultă că nu există deosebire, în cele din urmă, între a aserta şi a gândi, pe de o parte, şi *a face zgomot* (*sau a produce imagini mentale*) pe de altă parte. Dar asta înseamnă că (într-o asemenea concepţie) nu sunt o fiinţă *gânditoare*, ci doar un animal. Susţinerea unui asemenea punct de vedere reprezintă un fel de sinucidere.

Pentru a vedea dacă Wittgenstein are dreptate, să urmărim, deşi Wittgenstein nu o face, cum încearcă relativiştii să traseze distincţia între a avea dreptate şi a crede că ai dreptate.

Relativistul ar putea împrumuta ideea că adevărul este o *idealizare* a acceptabilităţii raţionale. El ar putea susţine că *X* este adevărat dacă '*X* este justificat pentru mine' *ar* fi înzestrat cu adevăr când îl observ, când raţionez asupra lui suficient de mult timp etc. Dar condiţionalii subiectivi de forma 'Dacă eu *aş*..., atunci aş crede cutare sau cutare' sunt, la fel cu orice judecată, interpretaţi diferit de diverşi filosofi.

Un realist metafizic poate privi judecăţile despre *ce s-ar întâmpla dacă* adevărate în sine sau false în sine într-un sens absolut, independent dacă vom fi vreodată justificaţi în acceptarea sau respingerea lor. Dacă relativistul interpretează judecăţile despre ceea ce ar crede sub anumite condiţii într-un mod realist, atunci el recunoaşte existenţa *unei* clase de adevăruri absolute, încât ar înceta să mai fie relativist.

Un non-realist sau un realist 'intern' priveşte judecăţile condiţionale ca judecăţi pe care le înţelegem (precum alte judecăţi) în mare parte prin surprinderea condiţiilor de *justificare*. Aceasta nu înseamnă că realistul intern *abandonează* distincţia dintre adevăr şi justificare, ci consideră adevărul (idealizare a *justificării*) drept ceva surprins, aşa cum surprindem orice alt concept pe calea (în bună parte implicită) înţelegerii factorilor care fac acceptabil raţional să susţinem ceva ca adevărat. Oare interpretul relativist al judecăţilor ar putea fi un realist intern?

Să ne reamintim că poziţia non-realistă pe care am descris-o deja (Capitolul 3) presupune o noţiune *obiectivă* de acceptabilitate raţională. Non-realistul respinge că adevărul este corespondenţă cu o lume deja constituită.

Tocmai asta îl face *non*-realist-metafizic. Dar respingerea teoriei metafizice a adevărului corespondenţă nu reprezintă acelaşi lucru cu a privi adevărul sau acceptabilitatea raţională drept ceva subiectiv. Nelson Goodman, care consideră adevărul şi acceptabilitatea raţională ca o specie a unui predicat mai general, 'a fi corect', aplicabil unui spectru larg, de la lucrările de artă la judecăţi, spune succint:

> Pe scurt, adevărul judecăţilor şi corectitudinea descrierilor, reprezentărilor, exemplificărilor, expresiilor – desenelor, schiţelor, dicţiei, ritmului – este mai întâi o problemă de ajustare la referinţă sau la alte aspecte, ori la modurile de organizare. Diferenţa între ajustarea unei perspective la o lume, a unei lumi la o perspectivă şi a unei perspective la o altă perspectivă dispare când se recunoaşte rolul jucat de perspectiva însăşi în acest proces. A cunoaşte şi a înţelege trec dincolo dobândirea convingerilor adevărate spre descoperirea modului în care perspectivele şi lumea se adaptează una la alta.

Scopul relativismului, caracteristica sa definitorie este, totuşi, să *respingă* existenţa unei noţiuni inteligibile de 'corespondenţă' *obiectivă*. Astfel, relativistul nu poate înţelege adevărul în termeni de condiţii *obiective* de justificare.

Tentativa de a folosi condiţionalii pentru a explica distincţia între *a avea dreptate* şi *a crede că ai dreptate* eşuează deoarece relativistul nu are o noţiune obiectivă pentru 'a avea dreptate' în legătură cu aceşti condiţionali mai mult decât în legătură cu orice alt tip de judecăţi.

În sfârşit, dacă relativistul de astăzi, ca şi vechiul Protagoras, decide să spună că nu există nici o diferenţă între 'Eu am dreptate' şi 'Eu cred că am dreptate' – că nu poate fi trasată o distincţie între a fi justificat şi a crede că eşti justificat – atunci, despre ce se vorbeşte într-o asemenea

concepţie – în afară de sunetele produse în speranţa că cineva va *simţi* despre ce este vorba? Ce este *cugetarea* dincolo de a făuri imagini şi a vorbi în minte în speranţa trezirii sentimentului de a avea dreptate? Relativistul ajunge să nege că vreun gând este despre ceva, atât în sens realist cât şi non-realist, deoarece nu poate distinge între *convingerea* că gândurile sale sunt despre ceva şi *faptul* că gândurile sale sunt despre ceva. Pe scurt, relativistul nu vede că existenţa unei forme obiective de 'a avea dreptate' este o presupoziţie proprie gândirii.

Există o relaţie interesantă între argumentul pe care l-am analizat (Platon–Wittgenstein) şi argumentul împotriva incomensurabilităţii pe care l-am atribuit lui Quine şi Davidson. Ultimii argumentează, de fapt, că un relativist consistent nu ar trebui să recunoască existenţa altor vorbitori (sau cugetători) în nici un fel (dacă 'zgomotele' lor sunt atât de incomensurabile, ele rămân numai zgomote), în vreme ce argumentul lui Platon şi Wittgenstein susţine că un relativist consistent nu poate considera un vorbitor ca un cugetător.

Ce ar fi de făcut?

Argumentele de mai sus m-au convins că două dintre cele mai cunoscute filosofii ale ştiinţei din acest secol sunt incoerente. (Mai mult, nici una dintre ele nu se limitează la o filosofie a ştiinţei). Aceasta mă conduce la noi reflecţii. Cum s-au impus asemenea puncte de vedere?

Pozitivismul logic era continuatorul, deşi diferit, al pozitivismului lui Mach. Pozitivismul lui Mach, sau empiriocriticismul, era, de fapt, o reluare a empirismului lui Hume într-un dialect diferit. Strălucirea lui Mach,

stilul său dogmatic şi entuziast, ca şi eminenţa ştiinţifică, făcea din empirismul său o adevărată mişcare culturală (Lenin se temea că bolşevicii s-ar putea converti la 'empiriocriticism' aşa că a scris o lucrare polemică la adresa lui.) Einstein, a cărui interpretare a teoriei relativităţii restrânse era operaţionalistă în spirit (în opoziţie cu interpretarea dată teoriei relativităţii generale) ştia că critica noţiunii de simultaneitate datorează mult lui Hume şi Mach deşi, spre dezamăgirea sa, Mach a respins teoria relativităţii restrânse.

Dar evenimentul cel mai însemnat care a condus la apariţia pozitivismului logic a fost revoluţia din logica deductivă. Prin 1879, Frege a descoperit un algoritm, o metodă mecanică de demonstrare, cunoscută astăzi drept 'logica de ordinul doi'. Această metodă este *completă* pentru teoria elementară a deducţiei (logica de ordinul întâi). Faptul că se putea elabora un algoritm prin care să se dovedească *toate* formulele valide ale logicii de ordinul întâi – un algoritm care nu cere o analiză semnificativă cu contribuţia întregii psihologii umane – este un fapt remarcabil. Ea a întreţinut speranţa că şi pentru logica inductivă – care era 'metoda ştiinţifică' – s-ar putea proceda la fel, iar aceşti doi algoritmi – algoritmul pentru logica deductivă (care este incomplet pentru logica de ordin superior) şi algoritmul care trebuie descoperit pentru logica inductivă – ar descrie exhaustiv 'reconstrucţia raţională' nu numai pentru raţionalitatea ştiinţifică, ci pentru orice fel de raţionalitate care ar merita acest nume.

Când am început cariera mea la Universitatea Princeton l-am cunoscut pe Rudolf Carnap, care petrecea

câţiva ani la Institutul pentru Studii Avansate. Într-o după-masă memorabilă, Carnap mi-a spus cum a devenit filosof. Pe când era student la fizică, a luat contact cu logica la seminarul lui Frege. Textul era *Principia Mathematica* (ne imaginăm ce înseamnă să studiezi *Principia* lui Russell şi Whitehead cu *Frege*!). Carnap era fascinat atât de logica simbolică dar şi de teoria relativităţii. Astfel, s-a decis să scrie o teză privind formalizarea relativităţii restrânse în limbajul simbolic din *Principia*. Asta pentru că Departamentul de Fizică din Jena nu a acceptat solicitarea lui Carnap de a deveni filosof.

Astăzi, ca urmare a rezultatelor negative, inclusiv a unor consideraţii ale lui Nelson Goodman, a devenit clar că *nu se poate* crea o logică inductivă formală completă. Unele aspecte importante ale logicii inductive pot fi formalizate (deşi adecvarea formalizării lor este controversată), dar rămân necesare judecăţi pentru 'raţionalizare', dacă acestea sunt construite prin alegerea vocabularului (sau, mai precis, împărţirea vocabularului în predicate *proiectibile* şi *neproiectibile*). Astăzi nimeni nu crede că există o metodă ştiinţifică pur formală (despre aceasta vezi Capitolul 8).

Povestea lui Carnap vine în sprijinul ideii că succesul înregistrat în cazul logicii deductive a jucat un rol crucial. Dacă acest succes a inspirat poliferarea pozitivismului logic, oare nu cumva eşecul în tentativa de a formaliza logica inductivă, descoperirea faptului că nu există un algoritm pentru ştiinţa empirică, a condus la poliferarea 'anarhismului'?

Nu vreau să insist asupra acestei sugestii; în tot cazul, pot fi invocaţi şi alţi factori. În vreme ce Kuhn s-a menţinut în limite moderate, atât Feyerabend, cât şi Michel Foucault au ajuns la extrem. Există analogii cu politicul în intenţiile lor: atât Feyerabend, cât şi Foucault fac legătura între criteriile instituţionalizate de astăzi şi capitalism, exploatare, ba chiar, represiune sexuală. De bună seamă că există numeroase motive pentru care relativismul este atractiv astăzi, ideea că toate instituţiile şi tradiţiile sunt rele fiind una dintre ele.

Un alt motiv este un anumit *scientism*. Scientismul pozitivismului logic este exagerat; dar eu cred că există scientism şi în spatele relativismului. Teoria că raţionalitatea se confundă cu aspecte ale culturii locale nu este niciodată îmbrăţişată în totalitate de către gânditorii anarhişti, ci există o limită naturală a acestei tendinţe: teoria reducţionistă. Definirea raţionalităţii printr-un program pentru un computer ideal este inspirată de ştiinţele exacte; iar raţionalitatea definită prin normele culturale locale este inspirată de antropologie.

Nu voi discuta aici aşteptarea provocată de lingvistica lui Chomsky, că psihologia cognitivă va descoperi algoritmi *înnăscuţi* care definesc raţionalitatea. Eu consider că aceasta este o modă intelectuală care va dezamăgi, ca şi speranţa pozitivismului logic într-o logică inductivă.

O parte a filosofiei din vremea noastră a moştenit o componentă scientistă din secolul al XIX-lea, moştenire care afectează mai multe domenii ale vieţii intelectuale. Nu neg importanţa logicii sau importanţa studiilor formale în semantică şi teoria confirmării. Dar sunt

înclinat să cred că ele sunt mai degrabă periferice în raport cu filosofia şi că, atâta vreme cât suntem obsedaţi de formalizare, ne putem aştepta să cădem mereu în scientism. Ambele tipuri de scientism pe care le-am evidenţiat sunt tentative de a evita o descriere sănătoasă şi umană a domeniului raţiunii.

Fapt şi valoare

Înţeleasă într-un sens suficient de larg, problema faptelor şi valorilor priveşte pe oricine. Sub acest aspect diferă de numeroase alte probleme filosofice. Oamenii obişnuiţi nu simt că ar trebui să aibă o opinie despre ce este lumea reală şi cum ne apare ea, de pildă. Problemele filosofiei limbajului, ale epistemologiei şi chiar ale metafizicii pot avea un interes opţional pentru oameni. Dar problema faptelor şi valorilor impune o atitudine. Orice persoană care reflectează *trebuie* să aibă o opinie cu privire la ea (care poate sau nu să fie aceeaşi cu opinia lor noţională). Dacă problema faptelor şi valorilor impune o alegere, răspunsul că faptele şi valorile sunt total separate, că

dihotomia 'judecăţi de fapt sau judecăţi de valoare' este una absolută, are statutul unei instituţii culturale.

Această dihotomie este admisă ca instituţie culturală, de cele mai multe ori, fără a ţine seama de ceea ce spun filosofii şi fără a avea în vedere dacă este *corectă* sau nu. Chiar dacă aş reuşi să vă conving că dihotomia fapt-valoare nu are o bază raţională, că nu poate fi argumentată raţional, sau chiar dacă un filosof mai bun decât mine ar putea arăta acest lucru printr-un raţionament imbatabil (care însă nu există în filosofie), în momentul următor, când veţi ieşi în stradă sau la un cocktail sau într-o discuţie particulară, veţi găsi pe cineva care va întreba 'Asta este o judecată de fapt sau una de valoare?' Punctul de vedere că nu este vorba despre fapte atunci când apreciem dacă lucrurile sunt bune sau rele sau mai bune sau mai rele etc., a devenit *instituţionalizat*.

Strategia argumentului meu nu este nouă. Doresc să reabilitez o mişcare oarecum discreditată în dezbaterea privind faptele şi valorile, respectiv acea mişcare care consideră că distincţia amintită este eronată, deoarece judecăţile factuale şi practicile de cercetare ştiinţifică, în care avem încredere când decidem ce este şi ce nu este fapt, presupun valorile.

Dovada că o asemenea mişcare există este replica pe care o generează, respectiv replica faţă de teza că ştiinţa presupune valori. Apărătorii dihotomiei fapt-valoare sunt de acord că ştiinţa presupune unele valori, cum ar fi valorile de adevăr, dar acestea nu sunt valori *etice*. Să imaginăm un oponent care adoptă teza că ştiinţa presupune o singură valoare, *adevărul*.

După cum am văzut, adevărul nu este o noţiune simplă. Ideea că adevărul este o copie pasivă a ceea ce este real (independent de mental şi de discurs) s-a prăbuşit sub

critica lui Kant, Wittgenstein şi a altor filosofi, chiar dacă mai găseşte susţinere.

Unii filosofi au făcut apel la *principiul echivalenţei*, potrivit căruia *a spune că o judecată este adevărată este acelaşi lucru cu a aserta acea judecată*, argumentând astfel că nu există o problemă filosofică a adevărului. Alţii apelează la lucrările lui Alfred Tarski, logicianul care arată cum, fiind dat un limbaj formalizat (o notaţie formală pentru a exprima anumite judecăţi, folosind logica simbolică) se poate defini 'adevărul' *relativ la acel limbaj* într-un limbaj mai înalt (aşa-numitul 'metalimbaj')[1].

Lucrarea lui Tarski era bazată pe principiul echivalenţei: de fapt, criteriul său pentru o definiţie corectă a adevărului era de a produce toate propoziţiile de forma *'P' este adevărată dacă şi numai dacă P*, ca în exemplul:

> (T) 'Zăpada este albă' este adevărată dacă şi numai dacă zăpada este albă.

ca teoreme ale metalimbajului (unde *P* este o propoziţie, potrivit regulilor de notare).

Dar principiul de echivalenţă este neutru filosofic şi aşa rămâne în lucrarea lui Tarski. Pentru *orice* teorie a adevărului, 'Zăpada este albă' este echivalentă cu '"Zăpada este albă" este adevărată'.

Filosofii pozitivişti ar răspunde că, dacă cunoaştem (T) de mai sus, *ştim* ce înseamnă '"Zăpada este albă" este adevărată' – înseamnă că *zăpada este albă*. Iar dacă nu înţelegem cuvântul 'zăpadă' sau 'albă', ar adăuga ei, atunci suntem în încurcătură! Dar problema nu este că nu înţelegem 'Zăpada este albă'; problema este că noi nu

[1] Pentru o prezentare accesibilă a tezelor lui Tarski, vezi *Meaning and the Moral Sciences*, Partea I, Conferinţa I.

înţelegem *ce înseamnă a înţelege* 'Zăpada este albă'. *Aceasta* este problema filosofică. Despre aceasta (T) nu spune nimic.

Oare acestea nu sunt în acord cu intuiţiile noastre despre adevăr? Dacă cineva ne abordează şi spune 'Nu vrei să cunoşti "adevărul"?' reacţia noastră este în general destul de bănuitoare pentru acea persoană din motivul că persoana respectivă nu ne spune nimic atât timp cât nu ştim care sunt standardele de acceptabilitate raţională la care acesta aderă: modalitatea de a desfăşura o cercetare pe care o consideră raţională, standardele sale de obiectivitate, când consideră că este raţional să încheie o investigaţie, ce fundamente are în vedere pentru a înzestra un verdict sau altul cu raţiune cu privire la problemele de care ar putea fi interesat. Ca aplicaţie la cazul ştiinţei, a spune că ştiinţa 'caută să descopere adevărul' este o judecată pur formală. Cu asta nu spunem mai mult decât că oamenii de ştiinţă nu vor să aserteze că zăpada este albă dacă zăpada nu este albă şi nu vor să aserteze că prin fire circulă electroni dacă prin fire nu circulă electroni etc. Dar aceste judecăţi pur formale sunt fără conţinut câtă vreme nu avem nici o idee despre sistemul de criterii pentru acceptabilitatea raţională care deosebesc metodele ştiinţifice de a determina dacă zăpada este albă de alte căi de a determina dacă zăpada este albă, ori metoda ştiinţifică de a determina dacă prin fire se deplasează electroni, de alte căi de a face acelaşi lucru etc.

Dacă noţiunea de comparare a sistemului nostru de convingeri cu realitatea neconceptualizată pentru a vedea dacă ele se potrivesc nu are sens, atunci afirmaţia că ştiinţa urmăreşte să descopere adevărul poate să fie acelaşi lucru cu faptul că ştiinţa încearcă să construiască un tablou al lumii care, în limite ideale, satisface un anumit criteriu de acceptabilitate raţională. Că ştiinţa urmăreşte să con-

struiască un tablou al lumii care să fie adevărat, este o judecată adevărată aproape vidă şi formală; scopul ştiinţei este surprins numai prin criteriile de acceptabilitate raţională implicite demersului ştiinţific. Pe scurt, răspunsul la teza susţinută de oponent că singurul scop al ştiinţei este să descopere adevărul (pe lângă faptul că ştiinţa are şi alte scopuri) este că *adevărul nu este scopul ultim*: dincolo de adevăr se află criteriile de acceptabilitate raţională, acestea fiind cele care constituie valorile implicite în demersul ştiinţific.

Ipoteza faţă de care tocmai mi-am manifestat dezacordul este cea formulată în Capitolul 1, că suntem creiere în container. Am argumentat că o asemenea ipoteză nu poate fi adevărată; dar să presupunem că există unii care nu au fost convinşi de argumentele aduse (ceea ce este probabil, căci argumentele filosofice nu conving niciodată pe toţi oamenii). Pe scurt, cele de mai sus se subsumează ipotezei că lumea este o halucinaţie colectivă.

Desigur, dacă lumea ar fi o halucinaţie colectivă, sunt mulţi oameni care nu şi-ar da seama de acest fapt. Pentru îndrăgostiţi, de pildă, diferenţa ar fi inexistentă sau foarte mică[2]. Ne imaginăm că ar fi acelaşi lucru şi pentru economişti. (De ce ar trebui ca un economist să fie încurcat de faptul că banii nu au realitate fizică? Unii dintre ei nu au realitate fizică pentru nici o teorie!)

Vreau ca cititorul să-şi imagineze că această teorie nebunească (şi aş afirma, incoerentă), teoria că toate creierele sunt în container, este susţinută nu de vreun lunatic izolat, ci de aproape toţi locuitorii dintr-un ţinut vast, cum ar fi Australia. Să ne imaginăm că în Australia, numai o mică minoritate crede aşa cum credem noi, în timp ce marea majoritate crede că suntem creiere în container. Poate

[2] Dar nu exclud alte posibilităţi.

australienii sunt discipoli ai unui guru, să-i spunem 'guru din Sydney'. Poate, atunci când vorbesc despre el, spun 'Oh, dacă ai vorbi cu guru din Sydney şi ai privi în ochii lui şi ai vedea cât este de bun şi de înţelept ai fi convins la rândul tău'. Dacă îi întrebi: 'Dar cum ştie guru din Sydney că suntem creiere în container dacă iluzia este atât de perfectă?', ei pot răspunde: 'Oh, guru din Sydney *ştie pur şi simplu*'.

Cum am arătat anterior, aici nu găsim un dezacord ştiinţific în sensul obişnuit. Ne putem imagina că australienii sunt tot aşa de buni ca şi noi în a face predicţii sau în a construi poduri rezistente (sau care par rezistente) etc. Ei ar putea chiar să accepte ultimele descoperiri ştiinţifice nu ca adevărate, ci drept descrieri corecte a ceea ce pare să fie. Am putea să ne imaginăm că nu sunt de acord cu noi cu privire la unele predicţii despre viitorul îndepărtat (de pildă, s-ar putea aştepta ca într-o zi, maşina automată să se strice şi oamenii vor începe să aibă halucinaţii colective încât va deveni evident că punctul lor de vedere este corect)[3] dar argumentul meu nu este afectat indiferent dacă ei fac asemenea predicţii, ori dacă nu se angajează în predicţii diferite de ale noastre, permise de teoria standard. Nucleul acestei situaţii imaginare îl constituie numărul mare de oameni care au convingeri în dezacord esenţial faţă de ale noastre.

Aici nu este vorba de un dezacord privind valorile 'etice'; australienii pot avea o etică similară cu a noastră. (Deşi grecii vechi ar spune că şi *înţelepciunea* este o valoare etică; iudaismul şi creştinismul au, în fapt, o etică foarte apropiată, bazată pe noţiunea de mântuire.)

---

[3] Dacă aceştia fac asemenea predicţii, atunci perspectiva lor nu mai este incoerentă în modul descris în Capitolul 1, afirmaţiile lor ar putea fi justificate (eventual) şi nu ar fi nevoie de ipoteza adevărului 'transcendent' (sau independent de justificare) pentru a fi înţelese.

Primul lucru pe care doresc să-l observ legat de ipoteticii australieni este că viziunea lor asupra lumii este *nebunească*. Uneori, 'nebunesc', este folosit aproape ca un termen pentru aprobare; dar aici nu am în vedere acest înţeles. Cred că am simţi o mare compasiune pentru oamenii care ar privi lumea într-un asemenea fel. Australienii ar fi consideraţi ţicniţi în sensul că au minţile *bolnave*; iar caracterizarea minţilor lor ca bolnave este una etică sau apropiată de etică. Dar cum altfel decât spunându-le adevărul în faţă ar putea cineva să-i convingă pe australieni? (sau să încerce să-i convingă, deoarece presupun că nu pot fi convinşi).

Un prim argument care poate fi adus este *incoerenţa* perspectivei lor. Am discutat despre aceasta în Capitolul 1. Aici este prezentă o incoerenţă *adâncă*, pentru care este nevoie de o argumentare filosofică (deci plină de controverse). Dar perspectiva australiană este incoerentă şi în alt mod. Unul dintre lucrurile pe care le am în vedere este de a determina căile prin care aflăm dacă o propoziţie este adevărată. Mai întâi încercăm să determinăm adevărul pe baza unei teorii cauzale a percepţiei, prin abordarea, prevăzută de teorie, a modului în care percepţiile rezultă din interacţiunea organelor noastre cu lumea externă. În al doilea rând, putem apela la o teorie a statisticii şi proiectării experimentelor astfel încât să putem arăta cum anume metodele pe care le folosim nu generează erori experimentale. Pe scurt, este elaborată o constrângere extrem de folositoare, ca teoria pe care o adoptăm asupra lumii să includă procedeele prin care stabilim dacă această teorie este corectă.

Sistemul australienilor nu are această proprietate de coerenţă (cel puţin dacă pentru coerenţă nu există un algoritm, ci ceva judecat printr-un sentiment de disconfort). Australienii sunt prizonierii unei iluzii atât de

perfecte încât nu este nici o cale raţională prin care guru din Sydney ar putea să ştie că sistemul de convingeri pe care l-a adoptat şi i-a convins şi pe ceilalţi să îl adopte este corect. Judecat după standardele noastre de coerenţă, sistemul *lor* de convingeri este incoerent.

Virtuţile noastre metodologice lipsesc sistemului lor de convingeri. Sistemul lor de convingeri, după cum a fost descris, acceptă că legile naturii sunt *imaginare*; dar oare un asemenea sistem ne spune dacă legile naturii care apar în imaginaţie sunt legile naturii care se aplică în afara containerului? Dacă nu se întâmplă aşa, atunci acestui sistem îi lipseşte un anumit tip de înţelegere deoarece acesta nu ne spune, chiar în propriii termeni, care sunt, în ultimă instanţă, legile naturii. Cu siguranţă, este violat briciul lui Occam. Este dificil sau imposibil ca briciul lui Occam să poată fi surprins printr-un algoritm, dar faptul că teoria creierelor în container postulează tot felul de obiecte în afara containerului care nu joacă nici un rol în explicarea propriilor experienţe, conform teoriei înseşi, este un caz în care este violată maxima '...nu multiplica entităţile fără să fie necesar'. Să numim o teorie care se supune briciului lui Occam *funcţional simplă.*

Până acum am văzut că metodele prin care decidem acceptabilitatea unei teorii ştiinţifice ţin seama dacă teoria are anumite 'virtuţi'. Presupun că metoda prin care a fost construită teoria nu poate fi analizată corect printr-un procedeu de verificare *propoziţie cu propoziţie.* Presupun că verificarea în ştiinţă este de natură holistă, adică sistemele teoretice ca întreg se confruntă cu testele empirice şi judecata despre cât de bine trec aceste teste este, în ultimă instanţă, o problemă de intuiţie care nu poate fi formalizată, aşa cum nu poate fi formalizată psihologia

umană în întregul său. Dar să revenim la problema inițială. Care sunt valorile implicite ale științei?

Am argumentat că, dacă ținem seama de valorile la care apelăm în critica teoriei creierelor în container cât și de alte valori care nu sunt prezente în acest caz, cum ar fi dezideratul de eficiență instrumentală pe care presupunem că îl împărtășim cu adepții creierelor în container, ajungem la o imagine a științei care presupune un sistem bogat de valori. De fapt, dacă luăm în considerare idealul de acceptabilitate rațională care este pus în evidență cercetând cum sunt teoriile pe care oamenii de știință și chiar ceilalți oameni le consideră acceptabile în mod rațional, atunci vedem că prin știință se încearcă realizarea unei reprezentări a lumii care întrunește caracteristicile de a fi eficientă instrumental, coerentă, comprehensivă și simplă funcțional. Dar de ce?

Motivul pentru care urmărim o asemenea reprezentare, și nu varianta 'bolnavă' a lumii noționale a australienilor adepți ai teoriei creierelor în container, este că a avea un tip de sistem reprezentațional constituie o *parte a înfloririi cunoașterii umane*, adică o parte a umanității depline sau a *eudaemoniei*.

De bună seamă, dacă realismul metafizic are dreptate și am admite că scopul științei este 'potrivirea' lumii noționale cu lumea în sine atunci s-ar putea contesta interesul pentru coerență, comprehensiune, simplitate funcțională și eficiență instrumentală numai pentru că acestea sunt instrumente care rezultă din această potrivire. Dar noțiunea de potrivire rațională între reprezentare și lume nu are sens. Prin negarea faptului că dorim un asemenea tip de potrivire metafizică cu lumea noumenală nu se neagă că dorim o corespondență empirică obișnuită (judecată după criteriile de acceptabilitate rațională) cu

lumea *empirică*. Dar lumea empirică, fiind opusă lumii noumenale, depinde de criteriile noastre de acceptabilitate raţională (şi, desigur, viceversa). Folosim criteriile noastre de acceptabilitate raţională pentru a alcătui un tablou teoretic al lumii empirice, iar apoi acest tablou serveşte pentru reevaluarea criteriilor de acceptabilitate raţională, după care ciclul se reia mereu. Dependenţa alcătuirii tabloului asupra lumii de metodă este subiectul altor cărţi ale mele; aici aş dori să subliniez cealaltă faţă a dependenţei, respectiv dependenţa criteriilor de acceptabilitate raţională faţă de tabloul empiric al lumii. Trebuie să avem criterii de acceptabilitate raţională potrivite cu tabloul empiric al lumii, care pun în evidenţă o parte din noţiunea inteligenţei speculative optime. Pe scurt, afirm că lumea reală depinde de valorile noastre (şi, din nou, viceversa).

Unele valori trebuie să fie obiective

Faptul că ştiinţa nu este 'neutră valoric' nu arată că valorile etice ar fi obiective sau că etica ar putea fi o ştiinţă. De fapt, nu există vreo 'ştiinţă a eticii' nici ca disciplină de laborator, nici ca disciplină deductivă. Cu mult timp în urmă, Aristotel afirma:

> Trebuie deci să ne mulţumim cu faptul că acei ce tratează despre astfel de lucruri şi pornesc de la ele înfăţişează adevărul într-o manieră nediferenţiată şi succintă şi că, vorbind despre şi pornind de la fapte atât de generale, ajung la concluzii de aceeaşi natură. În acelaşi mod trebuie privite şi cele spuse de noi aici. Aşadar, omul instruit caută în fiecare gen doar gradul de precizie implicat de natura subiectului; căci a pretinde unui matematician argumente persuasive ar fi ca şi când ai pretinde demonstraţii abstracte

unui orator.[4] (Aristotel: *Etica nicomahică*, Editura Ştiinţifică şi Enciclopedică, Bucureşti, 1988, p. 9, trad. Stella Petecel – *n. trad.*)

Dar faptul că acceptabilitatea raţională în cazul ştiinţelor exacte (care sunt exemplare pentru gândirea raţională) depinde de unele *virtuţi* cognitve cum ar fi 'coerenţa' şi 'simplitatea funcţională' arată că măcar unii termeni-valoare stau pentru proprietăţi ale lucrurilor pentru care se aplică şi nu numai în legătură cu sentimentele persoanei care foloseşte termenii respectivi.

Dacă termenii 'coerent' şi 'simplu' nu stau pentru *proprietăţile* unei teorii, chiar dacă imprecis sau fără a fi bine definite, ci numai pentru atitudinile pe care oamenii le au cu privire la teorii, atunci termenii folosiţi pentru acceptabilitate raţională, precum 'justificat', 'bine confirmat', 'cea mai bună explicaţie valabilă', ar trebui să fie, la rândul lor, subiectivi: deoarece acceptabilitatea raţională nu poate fi altceva decât parametrii de care depinde. Dar, aşa cum am argumentat în capitolul anterior, punctul de vedere că acceptabilitatea raţională este subiectivă se autorespinge. Astfel, suntem împinşi spre concluzia că măcar *aceşti* termeni-valoare au un anumit tip de aplicare obiectivă, un anume fel de condiţii obiective de justificare.

De bună seamă, s-ar putea încerca evitarea concluziei că există valori obiective, respingând teza că termeni precum 'coerent', 'simplu', 'justificat' ar reprezenta valori. S-ar putea susţine că aceştia stau pentru proprietăţi pe care le valorizăm, dar în mod subiectiv. O asemenea variantă întâmpină mai multe dificultăţi. 'Coerent' şi 'simplu' au prea multe trăsături comune cu termenii pentru valorile paradigmatice. Precum 'grijuliu', 'frumos' şi 'bun', 'coerent' şi 'simplu' sunt folosiţi adesea ca termeni laudativi. Concepţiile noastre despre

[4] *Ethica Nicomachea*, Cartea I, Cap. 3.

coerenţă, simplitate şi justificare sunt condiţionate istoric, la fel cu concepţiile despre 'grijă', 'frumuseţe' sau 'bunătate'; aceşti termeni epistemici figurează într-un anume tip de controverse filosofice perpetue aşa cum se întâmplă cu termenii pentru valori etice sau estetice. Concepţia despre raţionalitate a lui John Newman este clar diferită de aceea a lui Rudolf Carnap. Fiecare este suficient de convingătoare; ele pot fiinţa în acelaşi timp. Problema: *care este concepţia raţională despre raţionalitate* este dificilă în exact acelaşi fel cu justificarea dificultăţilor sistemelor etice. Nu există o concepţie *neutră* despre raţionalitate la care să apelăm.

S-ar putea încerca numeroase convenţii, cum ar fi, să spunem că 'justificat$_{carnap}$' este o proprietate diferită faţă de 'justificat$_{newman}$' şi că o 'judecată subiectivă privind valoarea' este implicată în decizia asupra ce înseamnă 'justificat$_{carnap}$' sau 'justificat$_{newman}$' pentru cuvântul 'justificat' dar că nici o judecată de valoare nu este implicată în stabilirea faptului că o judecată *S* este justificată$_{carnap}$ sau justificată$_{newman}$. Dar din punctul cui de vedere este folosit cuvântul 'fapt'? Dacă nu există o concepţie despre raţionalitate pe care cineva să o aibă în mod obiectiv, atunci noţiunea de 'fapt' este vidă. Fără valorile cognitive ale coerenţei, simplităţii şi eficienţei instrumentale nu avem nici o lume şi nici un fapt, nici măcar fapte despre ceea ce este relativ, deoarece acestea sunt în aceeaşi oală cu celelalte fapte. Aceste valori cognitive sunt considerate în mod arbitrar ca altceva decât o parte din concepţia holistă despre împlinirea umană. Lipsiţi de vechea idee realistă de adevăr corespondenţă şi de poziţia pozitivistă de justificare fixată prin criterii publice am rămas cu nevoia de a considera cercetarea pentru cele mai bune concepţii despre raţionalitate ca o activitate umană intenţionată care, ca orice activitate, este influenţată de înclinaţii sau obsesii şi este ghidată de ideea de bine.

Raţionalitatea în alte domenii

Dacă valorile implicite în ştiinţă, în special în ştiinţele exacte, arată o parte a ideii noastre de bine, cred că restul ideii de bine poate fi găsită dincolo de standardele acceptabilităţii raţionale în alte domenii ale cunoaşterii. La acest moment este nevoie să delimităm noţiunea de *standarde ale acceptabilităţii raţionale.*

Până acum am lucrat cu standardele acceptabilităţii raţionale în sens literal: standarde despre care spunem că ar trebui îndeplinite pentru a *accepta* judecăţi. Dar într-un sens mai larg, ele nu privesc numai modul în care judecăm adevărul sau falsitatea unui sistem de judecăţi, ci şi modul în care judecăm *adecvarea* lor şi claritatea. Există modalităţi – pur cognitive – altele decât falsitatea, în care un sistem de judecăţi nu reuşeşte să ofere o descriere satisfăcătoare.

Acest subiect atinge şi ştiinţele teoretice. Ştiinţele naturii nu descoperă numai judecăţi adevărate, ci şi judecăţi care sunt adevărate şi universale (legi), ca şi judecăţi care sunt adevărate şi *relevante.* Noţiunea de relevanţă aduce cu sine un spectru larg de interese şi valori. Astfel s-ar putea argumenta pentru teza că a cunoaşte lumea presupune valori şi nu pentru ca teza că lumea reală depinde de valorile noastre să devină mai radicală.

Când ajungem la raţionalitatea perceptuală, adică la standardele şi abilităţile implicite pe baza cărora stabilim dacă cineva este în măsură să ajungă la o abordare adevărată, adecvată sau clară chiar şi a celor mai simple fapte *perceptuale*, atunci ne dăm seama de numărul mare al factorilor care intervin. Recent, psihologii au ajuns la concluzia că şi în cele mai simple cazuri de percepţii avem de-a face cu amestecul

teoreticului. Aceasta este adevărat nu numai pentru nivelul neurofiziologic, ci şi pentru cel cultural. Un exponent al unei culturi care nu cunoaşte mobilierul nostru, s-ar putea afla într-o cameră pentru descrierea căreia, dacă nu ştie ce este o masă sau un scaun, va apela la elemente proprii culturii de unde provine. Descrierea ar consta în judecăţi adevărate care nu sunt adecvate.

Acest exemplu simplu arată că, pentru ca o descriere să fie adecvată, trebuie ca persoana care descrie să aibă un anumit set de concepte; aşteptăm de la persoanele raţionale ca în descrierile cu privire la un anume tip de lucruri să fie capabili să utilizeze concepte potrivite; faptul că acea persoană nu foloseşte un anumit concept poate fi temei pentru a-l critica atât pe el, cât şi descrierea lui.

Condiţiile descriptive cerute la un nivel simplu al discursului despre scaunele şi mesele dintr-o cameră fără oameni în ea sunt valabile şi în cazul discursului despre relaţiile interpersonale. Să considerăm termenii folosiţi în mod obişnuit pentru a descrie alte persoane, de pildă *politicos* sau *nepoliticos*. Aceşti termeni pot fi utilizaţi pentru a lăuda sau a blama; iar una dintre numeroasele distincţii care, dacă sunt uitate, conduc la confuzia asupra relaţiei dintre fapt şi valoare este distincţia dintre folosirea unei expresii lingvistice pentru a descrie şi folosirea ei pentru a lăuda sau blama. Dar o asemenea distincţie nu poate fi trasată numai pe baza vocabularului. Judecata că cineva este nepoliticos poate fi folosită pentru a blama; dar ar putea fi folosită şi pentru a descrie, ca şi pentru a explica sau a face predicţii.

De pildă, aş putea să-ţi spun: 'Nu-l lăsa pe John să-ţi rănească sentimentele. Pari să tragi concluzia din modul în care îţi vorbeşte că nu te place, dar asta este o impresie falsă. Nu are importanţă ce simte pentru tine atâta vreme cât se va comporta într-un mod în care sentimentele tale vor fi rănite.

El este mai degrabă un om nepoliticos, dar să nu crezi că asta are de-a face cu tine'.

În acest mic discurs imaginar cineva foloseşte cuvântul nepoliticos nu cu scopul de a-l blama pe John, ci cu intenţia de a prevedea sau explica modul de comportare a lui John faţă de altcineva. Atât predicţia, cât şi explicaţia pot fi *corecte*. În acelaşi fel, 'gelos' poate fi un termen de blam dar poate fi folosit şi fără intenţia de a blama. (Uneori ai tot dreptul să fii gelos).

Utilizarea cuvântului nepoliticos mi se pare un exemplu potrivit pentru modul în care distincţia fapt/valoare este vagă fără speranţă atât în lumea reală, cât şi în limbajul real. Importanţa unor termeni precum 'nepoliticos', 'obraznic', 'îndărătnic', 'nesuferit' etc. în evaluarea morală a fost arătată de Iris Murdoch în *Suveranitatea 'binelui'*[5]. Deşi fiecare dintre judecăţile 'John este un om foarte nepoliticos', 'John se gândeşte numai la sine', 'John ar face orice pentru bani' pot fi doar o simplă descriere în sensul cel mai pozitiv (să notăm că 'John ar face orice pentru bani' nu conţine termeni pentru valori), dacă cineva asertează conjuncţia acestor judecăţi trebuie să adauge 'John nu este un om bun'. Când credem că faptele şi valorile sunt independente credem că despre fapte se vorbeşte într-un dialect fizicalist sau birocratic, iar cu privire la valori trebuie folosiţi termeni abstracţi precum 'bine' sau 'rău'. Atunci când faptele însele sunt de tipul 'nepoliticos', 'se gândeşte numai la sine', 'ar face orice pentru bani', independenţa valorii faţă de fapte este mai dificil de susţinut.

Aşa cum am criticat o descriere care nu foloseşte concepte precum *masă* sau *scaun* când ar trebui, tot aşa

[5] Routledge and Kegan Paul, 1970.

trebuie criticat cineva care greşeşte în aprecieri precum 'politicos' sau 'spontan'; descrierea sa nu este adecvată.

Super-benthamiştii

Să modificăm argumentul anterior cu privire la teoria creierelor în container. De data aceasta să ne imaginăm că în Australia s-a dezvoltat o cultură care admite istoria, geografia şi ştiinţele noastre exacte, dar nu este de acord cu etica noastră. Nu ne vom opri la cazul familiar al super-naziştilor sau altul de acest fel, căci cred că este mai interesant cazul super-benthamiştilor. Să ne imaginăm Australia populată de oameni care au elaborat mijloace ştiinţifice de măsurare a 'gradului de plăcere' şi care cred că oamenii ar trebui să acţioneze pentru a maximiza plăcerea (cu cât gradul de plăcere este mai mare, cu atât îi corespunde un număr mai mare). Presupun că super-benthamiştii sunt extrem de sofisticaţi şi conştienţi de dificultatea de a prevedea viitorul şi de a estima consecinţele acţiunilor. Voi presupune, de asemenea, că sunt extrem de nemiloşi şi nu ar pedepsi acţiunile care ar duce la creşterea plăcerii, şi a măsurii acesteia, chiar dacă acestea presupun fapte oribile cum ar fi tortura copiilor sau condamnarea persoanelor nevinovate.

Desigur, pentru noi, o asemenea moralitate nu este de dorit. Cei mai mulţi dintre noi i-ar condamna pe super-benthamişti ca având un sistem de valori bolnav, ca fiind birocraţi fără scrupule sau cruzi. Ei sunt 'omul nou' în cea mai orbilă formă de manifestare. La rândul lor, ne-ar întoarce aceste invective spunând că suntem prea slabi, superstiţioşi, prizonieri ai unor tradiţii iraţionale etc.

Dezacordul dintre noi şi super-benthamişti este tocmai de tipul celui imaginat pentru a sublinia că două grupuri de oameni ar putea fi de acord asupra tuturor faptelor, mai

puţin asupra 'valorilor'. Dar să privim cazul mai îndeaproape. Orice super-benthamist este obişnuit cu faptul că, uneori, pentru a atinge cea mai mare satisfacţie sau cel mai mare număr, trebuie să mintă. Iar această faptă nu este condamnată dacă minciuna duce la maximizarea nivelului general de plăcere. Astfel, după un timp, folosirea expresiei 'sincer' printre super-benthamişti ar ajunge extrem de diferită faţă de folosirea acestui termen descriptiv de către noi. Acelaşi lucru s-ar întâmpla cu 'politicos' sau 'bun cetăţean' etc. Vocabularul utilizat de super-benthamişti pentru a descrie interacţiunile dintre oameni ar ajunge diferit de vocabularul valabil pentru noi. Nu numai că din el ar lipsi sau ar fi alteraţi dincolo de a putea fi recunoscuţi anumiţi termeni, dar ar inventa un jargon propriu (cu termeni exacţi pentru a descrie gradul de plăcere) care nu ar avea sens pentru noi. Structura lumii ar începe să se modifice. După un timp, super-benthamiştii vor ajunge să trăiască într-o lume diferită de a noastră.

Pe scurt, nu este adevărat că noi şi super-benthamiştii suntem 'de acord asupra faptelor, dar în dezacord privind valorile'. Pentru aproape toate situaţiile interpersonale descrierea faptelor pe care o dau ei va fi diferită de a noastră. Chiar dacă nici o judecată de-a lor despre o situaţie concretă nu este falsă, descrierea pe care o formulează nu este adecvată şi clară. Astfel, chiar dacă punem deoparte 'dezacordul cu privire la valori', nu am putea considera acceptabilă raţional reprezentarea asupra lumii pe care o propun super-benthamiştii. Asemănător adepţilor teoriei creierelor în container, inabilitatea de a da seama de felul în care este lumea este rezultatul direct al concepţiei *lor* bolnave privind împlinirea umană.

Subiectivismul despre bunătate

Am afirmat adesea că pasul de la 'John este politicos, sincer, blând, curajos, responsabil etc.' la 'John este bun din punct de vedere moral' presupune cel puţin o premisă nedemonstrată (şi nedemonstrabilă) şi anume, 'politeţea este moral bună'. Faptul că aceasta trebuie susţinută înainte de a trage concluzii morale din judecăţi 'factuale' arată că judecăţile etice nu sunt justificabile raţional.

Acest tablou al eticii ca un fel de piramidă întoarsă, cu un vârf (care nu are suport) constând în 'axiome etice' care susţin întregul corp de convingeri morale, credem că este naiv. Nimeni nu a reuşit să separe o structură axiomatică a eticii (cum a remarcat şi Aristotel în citatul de mai sus, maximele morale pe care să le putem enumera sunt întotdeauna adevărate numai 'în cele mai multe cazuri'). Un tablou de acest fel, al unui sistem de gândire sprijinit pe axiome fără întemeiere, a dat argumente scepticilor din orice domeniu. Scepticii care se îndoiesc de existenţa obiectelor materiale, de pildă, argumentează că principiul potrivit căruia 'dacă senzaţiile noastre sunt aşa cum ar fi dacă ar exista lume materială, atunci există lume materială' este o premisă fără întemeiere raţională dar care este invocată tacit ori de câte ori se afirmă că se 'observă' un obiect material sau se încearcă să se justifice convingerea în existenţa unuia. De fapt, etica şi matematica, ca şi discursul despre obiectele materiale, presupun concepte, nu axiome. Conceptele sunt folosite în observaţie şi generalizare şi sunt legitimate de succesul pe care îl avem atunci când le folosim pentru a descrie şi generaliza.

Un atac mai sofisticat asupra obiectivităţii eticii admite că convingerile noastre etice se sprijină pe

observarea unor cazuri specifice, pe 'intuiţii', maxime generale etc. şi nu pe o colecţie arbitrară de 'axiome', dar susţine că 'observaţia' etică este afectată de o boală incurabilă: *proiecţia.*

În conformitate cu această abordare, oamenii sunt în mod natural, chiar dacă nu perpetuu, miloşi. Astfel, când vedem că se întâmplă ceva teribil, de pildă cineva torturează un copil mic doar pentru plăcerea sa sadică, ne simţim (unii) dezgustaţi. Dar mecanismul psihologic al 'proiecţiei' conduce la experienţa de a simţi calitatea ca o calitate a actului în sine: spunem 'actul este dezgustător' când ar trebui să spunem 'reacţia mea a fost de dezgust'. Astfel, construim un corpus de aşa-numite 'observaţii etice', care sunt, de fapt, observaţii asupra propriilor noastre simţăminte subiective.

Povestea are şi forme mai complicate (ca oricare alta). Hume postulează o tendinţă umană spre 'simpatie' care devine mai extinsă sub influenţa culturii. Sociobiologii contemporani postulează un instinct pe care îl numesc 'altruism' şi vorbesc de 'genele altruismului'. Dar ideea de bază rămâne aceeaşi: există *simţăminte* etice, nu valori obiective.

Am văzut deja că nu este aşa: există cel puţin unele valori obiective, de pildă, *justificarea.* S-ar putea adăuga că valorile cognitive sunt obiective, în timp ce valorile etice sunt subiective; dar argumentul că nu pot exista valori obiective a fost respins.

Pentru a arăta ce este greşit în argumentarea în favoarea subiectivismului moral să ne reamintim argumentul folosit în Capitolul 2 împotriva realismului metafizic. Aceasta poate părea ciudat: subiectivismul nu se *opune* realismului metafizic? Pentru cei care cred astfel

orice argument împotriva realismului metafizic susţine subiectivismul; strategia pe care o voi adopta folosind acelaşi argument atât împotriva realismului metafizic, cât şi împotriva subiectivismului pare sortită eşecului.

Dar, în fapt, realismul metafizic şi subiectivismul nu sunt pur şi simplu 'opuse'. Astăzi tindem să fim prea realişti în domeniul fizicii şi prea subiectivişti în cel al eticii, iar aceste tendinţe au loc împreună. Asta se întâmplă deoarece vedem fizica (sau o fizică ipotetică viitoare) drept unica teorie adevărată şi nu ca o descriere acceptabilă raţional subjugată unor probleme şi scopuri, şi tindem spre subiectivism cu privire la descrierile pe care nu le putem reduce la fizică. A fi mai puţin realişti în ce priveşte fizica şi mai puţin subiectivişti asupra eticii sunt tendinţe strâns corelate.

Argumentul de la sfârşitul Capitolului 2 era îndreptat împotriva versiunii 'fizicaliste' sau naturaliste a realismului metafizic. Pentru a-l reaminti, să presupunem că interpretarea standard *I* (sub care 'pisică' se referă la pisici şi 'cireaşă' se referă la cireşe etc.) este sau coextensivă sau identică cu relaţia fizică *R*. Prin urmare *R* are loc între manifestări ale 'pisicii' (sau evenimente fizice pe care cineva le consideră astfel) şi pisici etc. Interpretarea non-standard *J* am considerat-o coextensivă cu o relaţie *R'* definibilă în termenii lui *R*, lumile posibile şi permutările folosite pentru a construi *J* (vezi *Anexa*). *R'* are loc între manifestări ale 'pisicii' (sau evenimente fizice pe care cineva le consideră astfel în cazul standard) şi *cireşe* etc. *R* şi *R'* sunt ambele 'corespondenţe': aceleaşi propoziţii sunt adevărate sub ambele corespondenţe. Acţiunile invocate pentru *R'*-adevărul unei propoziţii (acţiuni care au succes din perspectiva agentului) sunt aceleaşi cu acţiunile invocate când propoziţia este *R*-adevărată. Dacă *R* este

'identic cu referinţa'; dacă *R*, *R'* şi toate celelalte relaţii care circumscriu extensiuni pentru cuvintele noastre în modalităţi care satisfac constrângerile operaţionale şi teoretice nu sunt corecte în acelaşi fel; dacă *R* şi *R'*, dar şi celelalte relaţii de acest tip nu sunt la fel de corecte datorită uneia dintre ele – *R* – care este tocmai referinţa, atunci un asemenea fapt este inexplicabil din optica fizicalismului.

Argumentul nu este tocmai un contraargument la realismul metafizic (varianta fizicalistă), ci un argument împotriva *reducţionismului*. Dacă nu există nimic în tabloul fizicalist al lumii care să corespundă faptului clar că 'pisică' se referă la pisici şi nu la cireşe, atunci există un motiv decisiv pentru a respinge cerinţa ca toate noţiunile pe care le folosim trebuie reduse la termeni fizici, căci referinţa şi adevărul sunt noţiuni care nu pot fi reduse în mod consistent. Dacă eu cred că 'o pisică se află pe un covor' atunci mă angajez să cred că 'pisică' se referă la ceva (dar nu în acelaşi fel în care înţeleg realiştii metafizici referirea) şi că 'o pisică se află pe un covor' este adevărată (dar nu în acelaşi fel în care înţeleg realiştii metafizici adevărul).

Să vedem acum modul în care acest argument năruie temeiurile pentru subiectivismul etic. Teoria 'proiecţiei' ne dă o abordare a experienţei morale potrivit căreia experienţa morală este afectată de simţămintele subiective. Teoriei proiecţiei contrastează cu următoarea teză: 'toţi oamenii au, într-o anumită măsură, un simţ al justiţiei şi al ideii de bine. Astfel noi răspundem favorabil (cu întreruperi) la apeluri precum "fiţi buni cu străinul dintre voi, *deoarece ştiţi cum este să fii un străin în Egipt*". Simpatia noastră este uneori mărginită pentru că suntem convinşi că *se cuvine* să fie mărginită; noi simţim că o atrocitate este greşită (uneori) chiar atunci când nu avem nici o simpatie

faţă de victimă. Ajungem să stabilim similarităţi între injuriile suferite de alţii şi cele pe care le suferim noi înşine şi între beneficiile altora şi cele proprii. Inventăm cuvinte morale pentru trăsăturile morale ale situaţiilor şi ajungem, în mod gradual la generalizări morale explicite care duc la rafinarea noţiunilor noastre morale etc.'

Această abordare este, pe de o parte mai simplă, pe de alta mai complexă decât teoria proiecţiei (sprijinindu-se pe rolul argumentării în modelarea atitudinilor morale). Neîndoios, mulţi oameni inteligenţi simt că astăzi trebuie să respingem discursul despre 'simţul justiţiei' şi discursul despre 'a avea ideea de bine' (când acestea nu sunt luate într-un sens pur subiectiv), ca 'neştiinţifice'. În acest fel, cunoaşterea morală devine problematică; poate cu totul imposibilă.

Dar ce înseamnă aici 'neştiinţific'? Convingerea că există un lucru cum este dreptatea nu este o credinţă în *stafii*, nu este un 'simţ al dreptăţii' în sens paranormal care ne-ar permite să percepem asemenea stafii. Dreptatea nu este astfel încât cineva să propună să fie adăugată la lista lucrurilor recunoscute de fizică, aşa cum chimiştii din secolul al XVIII-lea adăugau 'flogisticul' la lista lucrurilor recunoscute de teoria chimică. Etica nu intră în conflict cu fizica aşa cum ar sugera termenul 'neştiinţific'; pur şi simplu 'drept', 'bun' şi 'simţul dreptăţii' sunt concepte ale unui discurs care nu este reductibil la discursul fizic. După cum tocmai am văzut, *alte* tipuri de discursuri esenţiale nu sunt reductibile la discursul fizic fără a fi ilegitime. A vorbi despre 'dreptate', ca şi a vorbi despre 'referinţă', poate fi *non*-ştiinţific, dar nu *ne*ştiinţific.

Pentru a face discursul mai explicit, să luăm un principiu fundamental al logicii sau al matematicii, bunăoară,

principiul că şirul numerelor întregi poate fi continuat la nesfârşit (orice număr are un succesor), sau principiul că o mulţime nevidă de numere conţine un cel mai mic număr printre elementele sale. Să ne oprim la următoarele afirmaţii: 'Aceste principii sunt adevărate pentru numere şi mulţimi de numere cu care avem de-a face în practică. De aceea ne apar ca necesare. Prin mecanismul numit „proiecţie", asociem acestor principii *simţământul de necesitate*; noi simţim că judecăţile au o "necesitate" misterioasă. În realitate, aceasta nu este o justificare. Din tot ce cunoaştem, aceste principii pot să nu fie adevărate'.

Virtual, nimeni nu ar fi de acord cu o asemenea viziune. Virtual, orice matematician ar spune ceva de genul: 'Cei mai mulţi oameni au intuiţii matematice pentru unele aspecte. Astfel, "vedem" intuitiv sau prin exemple (sau prin întrebări dibace precum sclavul din dialogul lui Platon) că principiile sunt adevăruri necesare'.

Kurt Gödel considera intuiţia matematică analogă *percepţiei*: obiectele matematice (pe care le numim 'concepte') sunt *dincolo*, iar intuiţia este în măsură să perceapă intelectual aceste entităţi platonice. Comparaţia făcută de Gödel între intuiţia matematică şi percepţie pune în evidenţă o formă simplă a percepţiei. Vederea nu ne oferă acces direct la lumea reală, ci o descriere a obiectelor care sunt parţial structurate şi constituite chiar prin vedere. Dacă luăm curcubeul fizicienilor drept curcubeul în sine, atunci curcubeul în sine nu are nici o bandă (analiza spectroscopică produce o distribuţie continuă a frecvenţelor); benzile roşu, portocaliu, galben, verde, albastru şi violet sunt trăsături ale curcubeului observat, nu ale celui fizic. Curcubeul perceput depinde de natura aparatului perceptual, de modul în care construim lumea pe cale vizuală în termenii lui Nelson Goodman. ('Obiectele' fizicienilor depind de modul în care este

structurată lumea, aşa cum reiese din varietatea fizicilor care construiesc 'acelaşi' obiect. Nu considerăm vederea eronată deoarece introduce benzi în curcubeu; cei care nu le-ar vedea ar avea defecte de vedere). Vederea este considerată bună când ne permite să vedem lumea 'aşa cum este' – adică lumea umană, funcţională, parţial creată de vedere.

Propunerea unei axiome a teoriei mulţimilor, cum este 'axioma alegerii', poate fi adoptată, parţial datorită acordului cu 'intuiţia' experţilor matematicieni şi parţial datorită consecinţelor sale. Dacă axioma alegerii nu duce la rezultate care reprezintă succese în optica matematicienilor, faptul că unii oameni o consideră 'intuitivă' are mai puţină importanţă. Intuiţia matematică este testată prin consecinţele pe care o are surprinderea principiilor matematice. Pe scurt, intuiţia matematică este bună când permite să vedem faptele matematice 'aşa cum sunt ele' – adică aşa cum sunt ele în lumea matematică în care sunt construite prin practica matematică umană (inclusiv aplicarea matematicii în alte domenii).

O descriere fiziologică sau psihologică a vederii nu poate spune dacă vederea benzilor de curcubeu are importanţă pentru vederea corectă. Cu atât mai puţin, un asemenea demers ar fi în măsură să spună dacă principiile la care ajunge intuiţia matematică sunt *adevărate*. Odată stabilite aceste lucruri, nu ne mai surprinde că descrierea proceselor cerebrale care au loc când cineva 'vede' că o acţiune nu este dreaptă nu poate spune dacă acţiunea este sau nu dreaptă.

Discursul despre 'percepţia' morală, ca şi discursul despre intuiţia matematică sau despre referire şi înţelegere, nu este reductibil la tabloul fizic al lumii. Asta nu înseamnă că fizica este 'incompletă'. Fizica poate fi completă relativ la scopurile fizicii. Incompletitudinea fizică se află la nivelul

teoriilor, imaginilor şi discursurilor, deoarece nici o teorie sau imagine nu este completă în raport cu toate scopurile. Dacă ireductibilitatea eticii la fizică arată că valorile sunt proiecţii, atunci culorile sunt, la rândul lor, proiecţii. La fel sunt numerele naturale. Prin urmare, aşa este întreaga lume fizică. Dar a fi o proiecţie în acest sens nu este acelaşi lucru cu a fi *subiectiv*[6].

Autoritarism şi pluralism

Am argumentat că avem nevoie de standarde de acceptabilitate raţională pentru a avea o lume; fie o lume a 'faptelor empirice', fie o lume a 'faptelor de valoare' (o lume în care există frumosul şi tragicul). Aşa ar trebui să se întâmple dacă am accepta că este imposibil să avem standarde de acceptabilitate raţională şi, în acelaşi timp, să le respingem sau să nu apelăm la ele. (Acest tip de scepticism care constă în respingerea unor standarde de acceptabilitate raţională are consecinţa dispariţiei conceptelor. Aşa cum a recunoscut Sextus Empiricus, ultima treaptă a scepticismului nu poate fi exprimată în

[6] O versiune neintenţionată a teoriei proiecţiei este citată de C.S. Lewis în *The abolition of Man* (Macmillan, 1947). Lewis are în vedere un text de liceu (pe care nu-l identifică). 'Vă reamintiţi că era vorba de doi turişti [*Lewis vorbeşte de binecunoscuta poveste despre Coleridge şi cascada*]: unul o numeşte „sublimă", iar celălalt „drăguţă"; Colleridge aprobă prima afirmaţie şi o respinge pe a doua cu dezgust. Gaius şi Titius [*numele date de Lewis pentru autorii neidentificaţi ai textului*] comentează astfel: „Când primul spune *Este sublim*, pare că face o apreciere cu privire la cascadă... De fapt, nu face o apreciere cu privire la cascadă, ci cu privire la propriile sale simţăminte. El spune de fapt, *simţămintele pe care le am sunt asociate cuvântului „sublim"*, pe scurt, eu am simţăminte sublime"'.

limbaj). Avem la fel de multă îndreptăţire să considerăm unele 'evoluţii' ale minţii ca bolnave (şi aşa facem) cu a privi unele tipare 'cognitive' ale minţii ca bolnave. Dar a spune acest lucru nu înseamnă respingerea pluralismului şi angajarea pe calea autoritarismului.

Dacă ne oprim la ştiinţă, teza că ştiinţa este o întreprindere obiectivă (printr-un standard de obiectivitate cu nuanţe antropocentrice, dar, aşa cum a remarcat David Wiggins, singurul standard de obiectivitate valabil pentru noi) nu înseamnă a susţine că orice problemă ştiinţifică are un răspuns determinat. Unele probleme ştiinţifice pot avea răspunsuri *obiectiv nedeterminate*, adică să nu existe nici o convergenţă cu privire la un răspuns chiar în limitele obiectivităţii cercetării ştiinţifice; iar unele probleme ştiinţifice pot avea răspunsuri determinate dar relative la context (de pildă, 'Care este cauza pentru care John a suferit un atac de cord?' poate avea răspunsuri diferite depinzând de cine întreabă şi de ce). Similar, teza că cercetarea etică este obiectivă, în sensul că unele 'judecăţi de valoare' sunt adevărate într-un mod definit, iar altele sunt false în mod definit, sau, mai general, că unele poziţii (şi unele 'ideologii') sunt greşite în mod definit iar unele sunt inferioare altora, nu este acelaşi lucru cu teza că nu există deloc cazuri nedeterminate. (O formă importantă de nedeterminare a fost pusă în evidenţă de Bernard Williams: când toate alternativele sunt atât de oribile încât nu rămâne nici o variantă care să fie aleasă de o persoană ideală sau înţeleaptă). Faptul că există contexte relative în etică este prea evident pentru a insista.

Principala diferenţă în raport cu Aristotel este un mai accentuat pluralism. Aristotel recunoştea că pentru tipuri diferite de indivizi ar putea fi adoptate şi cuprinse în constituţii forme diferite de *eudaemonia*. El credea că

idealul constă într-o constituţie prin care fiecare primeşte ce i se cuvine; că într-o lume ideală (dincolo de problemele obişnuite despre cine ar cultiva câmpul sau cine ar coace pâinea) oricine ar fi un filosof. Suntem de acord cu Aristotel că fiecărui individ îi corespunde o formă proprie de desăvârşire, dar credem că, chiar într-o lume ideală, ar exista constituţii diferite, diversitatea fiind o parte a idealului. Vedem unele grade ale tensiunii tragice între idealuri, adică îndeplinirea unor idealuri exclude îndeplinirea altora. Dar subliniez din nou că convingerea în pluralitatea idealelor nu înseamnă că orice ideal de împlinire umană este la fel de bun. Unele idealuri trebuie respinse ca bolnave, infantile, greşite, unilaterale etc.

Angajamentul faţă de obiectivitatea etică nu ar trebui confundat cu ceva foarte diferit, cum ar fi angajamentul faţă de autoritarismul etic sau moral. Poate că această confuzie i-a determinat pe unii filosofi[7] proeminenţi să respingă ceea ce priveau ca o versiune limitată de 'non-cognitivism' şi să spună 'Grija pentru "traiul cel mai împlinit" este a fiecărui om, autoritatea finală trebuie să fie individul însuşi'. (Să remarcăm ambiguitatea termenului 'autoritate finală': înseamnă autoritatea politică? înseamnă autoritate epistemologică? sau nu are importanţă, aşa cum sugerează 'non-cognitivismul'?) Considerarea persoanelor ca agenţi morali autonomi cere să fim de acord cu dreptul de a alege propriile standarde oricât de respingătoare mi s-ar părea aceste alegeri. Conform liberalismului politic, nici guvernul nu trebuie să se amestece în alegerile individuale prin impunerea religiei sau a unui standard de

[7] David Wiggins, 'Truth, Invention and the Meaning of Life', *Proceedings of British Academy*, vol. LXII, 1976.

moralitate. Dar opoziţia înverşunată pentru toate formele de autoritarism politic şi moral nu trebuie să ne angajeze la relativism moral şi la scepticism. Motivul pentru care amestecul guvernului în morala individuală este greşit nu este că nu are importanţă ce formă de viaţă adoptă individul. (Dacă nu ar exista eroare morală atunci nu ar fi eronat ca guvernul să impună alegerile morale.) Faptul că multor cercetători le este teamă că, dacă admit vreo formă de obiectivitate morală atunci justifică acţiunile guvernului şi se ajunge la o obiectivitate morală impusă, este fără îndoială unul din motivele pentru care atât de mulţi subscriu la subiectivismul moral.

Raţiune şi istorie

O dată cu succesul ştiinţei, s-a văzut că multe probleme nu pot fi abordate cu metodele ştiinţelor exacte, cele ideologice sau etice fiind un exemplu limpede. De aceea, simultan cu creşterea admiraţiei şi respectului pentru fizician, cosmolog, biolog a intervenit descreşterea respectului şi încrederii faţă de gânditorul politic, moralist, economist, muzician, psihiatru etc.

Unii s-au împăcat cu această situaţie culturală şi au susţinut că, într-adevăr, nu există cunoaştere dincolo de domeniul ştiinţelor exacte (iar ştiinţele sociale pot avea succes numai în măsura în care adoptă strategia ştiinţelor exacte). Acest punct de vedere poate îmbrăca forma pozitivismului şi materialismului, sau unele combinaţii ale lor.

Alţii au încercat să argumenteze că ştiinţa este subiectivă şi arbitrară – aceasta este interpretarea populară a lucrării de succes a lui Kuhn *Structura revoluţiilor ştiinţifice*, chiar dacă nu era şi intenţia lui Kuhn. Alţii – cum ar fi marxiştii şi filosofii religioşi – adoptă un fel de dublu standard lăsând problemele tehnice pe seama ştiinţelor exacte şi inginerilor, iar problemele etice şi cele ideologice pe seama altor instanţe, cum ar fi partidul, viitorul utopic sau biserica. Doar puţini se pot mulţumi cu aceste variante – cu scientismul extrem, indiferent că este în variantă pozitivistă sau materialistă, cu subiectivismul şi relativismul radical, sau cu un fel de măsură dublă. Merită să examinăm această problemă datorită disconfortului pe care îl provoacă.

Într-un fel, avem de-a face cu o *falsă* problemă. Aceeaşi persoană care argumentează că maximele etice şi opiniile politice nu sunt verificabile susţine cu pasiune opiniile *sale* etice şi politice; iar relativiştii fac acelaşi lucru cu privire la relativismul lor. De aici rezultă că nimeni nu poate trăi în mod consistent în relativism; astfel, relativismul ne împinge spre existenţialismul anului 1945 ('totul este absurd, dar trebuie să alegem'). Oare sunt atât de diferite?

Pentru a ne fixa ideile, să ne reamintim remarca unui filosof al secolului al XIX-lea al cărui utilitarism este o formă de relativism. Mă refer la Bentham şi la judecata sa provocatoare: 'lăsând prejudecăţile deoparte, jocul de pushpin* are aceeaşi valoare cu arta şi ştiinţa poetică şi muzicală'. *Dincolo de prejudecăţi, pushpin este la fel de bun ca şi poezia.*

Cititorul modern este şocat de diferenţa faţă de valorile noastre culturale. Arta a fost ridicată de noi pe un loc mult mai înalt decât pe vremea lui Platon sau în Evul Mediu. După cum au remarcat mai mulţi autori, pentru o

---

* (joc de copii cu două ace încrucişate – *n. trad.*)

anumită categorie de persoane educate, arta de astăzi este religie, este cea mai scurtă cale spre mântuire.

Teza lui Bentham este că preferinţa pentru 'artele şi ştiinţele muzicii şi poeziei' în raport cu un joc de copii este pur subiectivă, ca şi preferinţa pentru îngheţata cu cremă de vanilie sau cea cu cremă de ciocolată. El nu neagă că muzica şi poezia ar avea o valoare mai mare decât acel joc copilăresc; în contextul utilitarismului său, faptul că cei mai mulţi preferă muzica şi poezia conduce la o mai mare utilitate pentru muzică şi poezie, deci la o valoare mai ridicată a acestora. Dar valoarea este un produs al 'prejudecăţii' (al interesului subiectiv); nu există fapte care să ierarhizeze valoric pushpin şi poezia, în afară de acela că ultima este preferată de mai mulţi oameni. Nu preferăm poezia pentru că ar avea o valoare mai mare, ci, mai degrabă, poezia are o valoare mai mare pentru că oamenii o preferă. (Aparent fără nici o motivaţie).

Susţinerea unei asemenea teze este atât de precară, încât o face neplauzibilă. Să ne oprim asupra preferinţei 'subiective'.

Oamenii par a accepta următorul model pentru preferinţa subiectivă. Există *C* care are gust de îngheţată cu cremă de ciocolată; există *V* care are gust de îngheţată cu cremă de vanilie. Există două sentimente *L* şi *D* care sunt 'plăcere' şi 'neplăcere'. Iar atunci când lui Jones îi place şi lui Smith nu-i place vanilia (sau lui Smith îi place, iar lui Jones nu-i place ciocolata) admitem că Jones are experienţa *V*+*L* când mănâncă vanilie şi *C*+*D* când mănâncă ciocolată, în vreme ce Smith are experienţa *V*+*D* când mănâncă vanilie şi *C*+*L* când mănâncă ciocolată.

O asemenea abordare este, totuşi, naivă aşa cum a argumentat mai demult Köhler. Gustul de vanilie care îi place lui Jones nu este gustul care nu îi place lui Smith care

nu suportă îngheţata cu cremă de vanilie. Mai degrabă lucrurile stau astfel: să spunem că vanilia are pentru Smith calitatea V_s. V_s este un gust neplăcut; ne-am putea imagina că cineva ar avea experienţa V_s şi i-ar plăcea, dar numai parţial şi datorită unui fel de disociere şi represiune. Pe scurt, psihologic, dacă nu metafizic, V_s este 'intrinsec' neplăcut. Iar Smith simte D (gust neplăcut) când mănâncă vanilie deoarece vanilia are gustul V_s (pentru el). În mod analog, V_j, calitatea gustului de vanilie pentru Jones, este intrinsec 'plăcută' (aşa explicându-se de ce Jones simte L). În limbajul lui G. E. Moore, gustul V_j şi valoarea pozitivă formează o unitate organică pentru Jones, iar gustul V_s şi valoarea negativă sunt o unitate organică pentru Smith. Fenomenologic, ele nu pot fi separate aşa cum se întâmplă în notaţia V_s+D şi V_j+L. Aproape sigur (cu excepţia apariţiei unor factori de represiune şi disociere), lui Smith i-ar place îngheţata cu cremă de vanilie dacă ar simţi V_j, iar lui Jones nu i-ar place, dacă ar simţi V_s în loc de V_j.

De ce considerăm preferinţa pentru vanilie sau pentru ciocolată, subiectivă? Adică, de ce o considerăm subiectivă chiar dacă nu credem că *toate* judecăţile de valoare sunt subiective, sau, conform cu Bentham, că, dincolo de prejudecăţi, pushpin este la fel de bun ca poezia? De bună seamă, dacă credem că toate preferinţele sunt subiective, aşa o vom considera şi pe aceasta, dar problema interesantă este de ce această judecată nu ne apare obiectivă decât, poate, dacă am fi Jones sau Smith, de ce nu are tipul de obiectivitate pe care fără îndoială îl au multe judecăţi de valoare?

Nu există nici o nepotrivire. Dacă admitem că există judecăţi de valoare obiective (sau justificate) putem admite foarte uşor că unele dintre judecăţile aprins disputate sunt obiectiv corecte. Naziştii discutau în jurul judecăţii că

deoarece apartenența lor rasială este greșită, uciderea evreilor este de dorit, dar antinaziștii nu considerau 'subiectiv' dezacordul în această privință cu naziștii. Cei care cred că homosexualii ar trebui să aibă drepturi depline în societatea noastră sunt în dezacord violent cu cei care cred că activitatea homosexuală și drepturile civile ale homosexualilor ar trebui să fie prohibite; nici una dintre părți nu consideră preferința sa 'subiectivă'. Într-adevăr, prezența dezacordului îi face pe oameni să fie siguri că poziția lor morală este întemeiată. Astfel, caracterul subiectiv al dezacordului între Smith și Jones nu este generat de faptul că 'unii oameni preferă ciocolata și unii oameni preferă vanilia'.

Altă explicație posibilă este că pentru cei mai mulți oameni, preferința pentru ciocolată sau vanilie nu este foarte puternică și nici importantă. Dacă jumătate din populație nu ar putea să suporte ciocolata dar le-ar plăcea vanilia și cealaltă jumătate nu ar suporta vanilia în schimb le-ar plăcea ciocolata, încă am putea considera aceasta ca o 'problemă de gust', adică în mod subiectiv. Nici existența 'neutrilor' nu ar fi decisivă.

În opinia mea, preferințele lui Jones sau Smith *ar deveni* importante dacă ar fi corelate cu trăsături ale minții și caracterului din motive biologice sau psihice. Să imaginăm, un experiment mental în care o trăsătură de caracter este considerată bună, atât prin ea însăși, cât și datorită influenței sale asupra simțămintelor, judecății și acțiunii, iar o altă trăsătură de caracter este privită ca rea, atât în sine, cât și datorită efectelor sale și oricine știe că trăsătura bună se manifestă prin preferința pentru vanilie, iar cea rea prin preferința pentru ciocolată; în acest caz, prima preferință ar fi privită ca normală și corectă, iar a doua ca perversă sau monstruoasă sau ceva de acest fel.

Nu vreau să afirm că toate preferinţele trebuie judecate moral prin trăsături de caracter. Unele preferinţe sunt foarte importante în sine: cineva care crede că este minunat să torturezi copii mici pentru a te simţi bine ar fi condamnabil (dacă ar fi normal) pe baza unei asemenea atitudini. Dar dacă obiectul preferinţei nu este privit ca important, atunci analizarea preferinţei ca o problemă de gust depinde de ce anume pune în evidenţă acea preferinţă. Judecăţile de valoare sunt adesea grupate; iar gruparea judecăţilor de valoare exprimă trăsături durabile ale minţii, personalităţii şi caracterului. *Independenţa* judecăţii 'Eu prefer înghеţata cu cremă de vanilie celei cu cremă de ciocolată' faţă de orice complex interesant şi semnificant de judecăţi este tocmai ceea ce o face subiectivă (dacă nu este prezentă o alegere bazată pe motivaţii intrinseci).

Chiar dacă preferinţa lui Smith pentru vanilie este subiectivă, nu înseamnă că este iraţională sau arbitrară. Smith are o motivaţie pentru a-i plăcea vanilia – cea mai bună motivaţie posibilă – respectiv, *gustul pe care îl are*. Valorile pot fi 'subiective' când sunt relative, dar şi obiective; este obiectiv faptul că vanilia are un gust mai bun decât ciocolata pentru Smith. În *Suveranitatea binelui*, Iris Murdoch subliniază că filosofi atât de diferiţi între ei precum existenţialiştii francezi şi pozitiviştii logici împărtăşesc un model comun al judecăţilor de valoare, modelul raţiunii ca mijloc de aprovizionare a minţii cu fapte neutre pe baza cărora ar trebui să alegem în mod arbitrar 'valori' – alegerea valorilor trebuie să fie arbitrară deoarece 'faptele' sunt (prin definiţie) neutre. Dar când voinţa nu dă nici o motivare judecăţii (motivaţia numai suplineşte 'faptele', potrivit acestui model) nu există *raţiune* pentru alegerile arbitrare; iată de ce ultimul temei al alegerii morale este numit 'absurd' de către filosofii

francezi şi de ce filosofii cu înclinaţii mai naturaliste vorbesc de instinct sau emoţie (succesorii lui Bentham propun plăcere).

În cazul pe care l-am examinat, modelul existenţialist-pozitivist nu se potriveşte. 'Faptul' – gustul în sine – şi valoarea – bunătatea gustului – sunt unificate, cel puţin din perspectivă psihologică. Calităţile experimentate nu sunt, în general, neutre, aşa încât ele cer mereu răspunsuri şi atitudini. S-ar putea să nu se ţină seama de aceste cerinţe aşa cum un copil învaţă să suporte durerea unei injecţii datorită avantajului dat de creşterea imunităţii, dar *prima facie* bunătatea sau răul unei experienţe particulare pot fi numai cu greu negate. (Această opinie destul de interesantă era adoptată de Platon şi de medievali – noi suntem, probabil, prima cultură care considerăm experienţa neutră).

Non-neutralitatea experienţei este prezentă şi în cazul pushpin/poezie. Noi credem că este aproape imposibil să ne imaginăm că un ins care apreciază într-adevăr poezia, cineva care este capabil să deosebească poezia adevărată de simpla versificaţie, capabil să răspundă la marea poezie, *ar* prefera un joc copilăresc artelor care ne îmbogăţesc viaţa, precum poezia sau muzica. Noi *avem* un motiv să preferăm poezia faţă de pushpin, iar acest motiv constă în resimţirea experienţei marii poezii şi în efectele marii poezii – creşterea sensibilităţii şi imaginaţiei prin lărgirea repertoriului de imagini şi metafore, şi integrarea acestora cu percepţiile şi atitudinile obişnuite atunci când un poem trăieşte în noi un număr de ani. Iată experienţe care sunt bune *prima facie* – şi nu numai bune, ci înălţătoare.

Dacă acestea pot fi motivaţii suficiente pentru judecăţile de valoare – motivaţii care sunt bune pentru indivizi particulari care emit judecăţi de valoare

particulare, – nu înseamnă că toate judecăţile de valoare sunt raţionale. Judecăţile de valoare, judecăţi care pot trezi pasiunile celor care le împărtăşesc, în numele cărora unii oameni i-au ucis şi torturat pe alţii, au fost adesea temeiul pentru un amestec nesănătos de impulsuri agresive şi idei narcisiste. Nu este surprinzător că, atunci când un istoric/filosof relativist cum este Michel Foucault scrie despre trecut, se concentrează asupra ideilor iraţionale şi judecăţilor de valoare. Dar este important să elucidăm de ce procedează astfel.

Foucault scrie despre epoca modernă timpurie (secolele al XVI-lea şi al XVII-lea) şi despre ideologie şi cultură în general. Cunoştinţele sale asupra faptelor sunt de domeniul legendei aşa că mulţi specialişti contestă caracterul lor de 'fapte'. În vreme ce unele cărţi ale sale sunt abstracte (*Arheologia cunoaşterii*), altele sunt mai de grabă aplicate: *Naşterea clinicii* şi *Naşterea închisorii. Naşterea clinicii* este, probabil, cea mai bună lucrare a lui Foucault şi conţine o importantă parte din teoriile sale.

Foucault încearcă să arate că 'clinica', adică instituţia spitalului şi instituţiile medicale legate de aceasta, este urmarea apariţiei unei anumite ideologii cu privire la boală şi sănătate, tot atât de mult ca şi rezultatul creşterii cunoaşterii ştiinţifice şi tehnice. Această ideologie era conectată cu schimbările ideologice mai largi, îndeosebi cu dezvoltarea individualismului în secolul al XVII-lea. El sugerează că spitalul nu este un loc foarte bun pentru a trata pacienţii şi, mai mult, convingerea noastră că spitalul ajută la vindecare este un fel de prejudecată ideologică şi absurdă.

Sugestia mai extinsă care se desprinde este că convingerile ideologice şi judecăţile de valoare asociate sunt

mai degrabă arbitrare[1]. Nu există un mod obiectiv pentru a ne afla în domeniul ideologic (cu excepţii, de bună seamă, datorită punctului de vedere ciudat privind propria 'Arheologie a cunoaşterii', pretins obiectivă).

Pentru a vedea ce îl conduce pe Foucault, să considerăm un exemplu mai familiar şi mai puţin controversat. În Evul Mediu se credea că monarhia este forma de guvernământ cea mai potrivită şi naturală. Această convingere era bazată parţial pe convingeri factuale considerate neîntemeiate astăzi (că democraţia ar conduce inevitabil la anarhie şi tiranie) şi în parte pe autoritatea bisericii. La rândul ei, doctrina bisericii era bazată parţial pe consideraţii politice, dar acestea nu erau luate în seamă, deoarece biserica le considera inspirate şi urmare a interpretării cuvântului şi voinţei lui Dumnezeu. Ceea ce sugerează Foucault este că convingerile susţinute în trecutul recent şi, în consecinţă, convingerile pe care le susţinem astăzi nu sunt mai raţionale decât credinţa medievală în dreptul divin al regelui.

Să analizăm credinţa în dreptul divin al regelui. Dacă nu admitem că există *bune motivaţii* pentru a crede în existenţa unui Dumnezeu personal care impune o anumită formă de organizare şi structură socială, atunci această convingere va fi imediat considerată ca iraţională (ceea ce nu înseamnă că nu corespunde unor nevoi psihologice).

---

[1] Şi, în plus, gândirea noastră este determinată de limbajul folosit. Foucault vorbeşte de 'sistemele implicite care ne determină comportamentul cel mai familiar în mod inconştient' (vezi J. Simon, 'A Conversation with Michel Foucault', *Partisan Review*, Nr. 2, 1971, p. 201). Structuralismul francez, cel puţin prin reprezentanţi precum Foucault, Althusser, Lacan, Deleuze etc., pare caracterizat prin: (1) determinism; (2) relativism şi (3) afirmaţia că structuralismul este o 'disciplină lingvistică'.

Chiar dacă credem în Dumnezeu, dacă nu credem că biserica are acces la voinţa sa, vom considera credinţa în dreptul divin al regelui ca iraţională. În fine, chiar credincioşii catolici vor admite că sprijinul acordat de biserică monarhiei în Evul Mediu era mai degrabă rezultatul unor împrejurări politice decât al revelaţiei şi teologiei. Pe scurt, credinţei în dreptul divin al regelui îi lipseşte, şi i-a lipsit întotdeauna, o bază raţională.

În acest caz, cum a apărut o asemenea credinţă? Răspunsul obişnuit pune apariţia ei pe seama unor factori economici şi politici (nu trebuie să fim marxişti pentru a admite că asemenea factori au rolul lor în determinarea ideologiei), ca şi în seama unor factori psihologici. Confortul generat de credinţa în determinarea divină este limpede, aşa fiind şi confortul generat de credinţa în infailibilitatea bisericii şi în implicarea divinităţii privind ordinea socială. Pe scurt, această credinţă a fost determinată de un fel de automulţumire narcisistă şi de condiţiile sociale. Dacă avem de-a face cu o credinţă *tipică*, dacă este reprezentativă pentru toate convingerile ideologice, atunci asemenea factori determină toate ideologiile.

Deoarece acceptă un asemenea scenariu, mulţi gânditori francezi îi stimează pe Marx, Freud sau Nietzsche. Marx, Freud şi Nietzsche au ceva în comun: ei consideră ideile noastre etice şi religioase comune ca reflectări ale *iraţionalului*, ale intereselor de clasă (la Marx), ale inconştientului (la Freud şi Nietzsche) sau ale voinţei de putere (Nietzsche).[2] Dincolo de ce ne place să considerăm drept cele mai profunde, spirituale şi morale acţiuni, fierbe

[2] Nu îi acuz pe Marx, Freud sau Nietzsche că au tras concluzii relativiste de aici.

cazanul interesului economic şi fanteziilor egoiste. Aici se află nucleul relativismului de astăzi.

În acelaşi timp, relativiştii nu pot folosi termenul 'iraţional' în felul pe care l-am prezentat aici. Când am arătat aceste pagini unui prieten relativist el a fost indignat de consideraţiile mele la adresa dreptului divin al regelui. Nu mi-am dat seama că oameni inteligenţi erau convinşi prin argumente filosofice că această doctrină era corectă? Ceea ce ofeream era o explicaţie simplistă marxistă sau freudiană? De bună seamă că credinţa în dreptul divin al regelui era *raţională*!

Răspunsul meu a fost că poate exista un sens în care orice punct de vedere care este apărat cu inteligenţă şi convingător în faţa unor prejudecăţi culturale este raţional, dar acest sens nu poate fi unul important. Evreii l-au acceptat pe Moise ca profet şi legislator deoarece doctrina sa răspundea unor *nevoi* religioase, culturale şi sociale; aceasta nu este acelaşi lucru cu convingerea prin argumente raţionale. Mai târziu, profeţii investeau regii evrei (după ce au încercat să-i facă să se răzgândească pe evrei să aibă regi); asta nu este o dovadă că regii creştini de mai târziu erau investiţi de către divinitate. Creştinismul, care adoptase Biblia de la evrei, devine religia Imperiului Roman – fără ca populaţia sau împăratul să aibă dovezi raţionale asupra adevărului său. Împăraţii romani erau apoi învestiţi (la fel cu regii evrei); nici de aici nu rezultă că ei erau desemnaţi în mod divin. În consecinţă, după acceptarea presupoziţiilor creştinismului, se puteau aduce argumente raţionale pentru a deriva dreptul divin al regilor *din aceste presupoziţii*. Dar a exprima acest lucru prin a spune că în Imperiul Roman şi în Evul Mediu, 'credinţa în dreptul divin al regilor era pe deplin raţională' înseamnă o deviere de la noţiunea de raţionalitate.

Hegel, care a introdus ideea că raţiunea însăşi se schimbă pe parcursul istoriei, operează cu două noţiuni de raţionalitate: există un sens în care ceea ce este raţional este măsurat prin nivelul de dezvoltare a spiritului în procesul istoric la un moment dat; acesta este analog sensului potrivit căruia 'credinţa în dreptul divin al regelui era raţională la momentul respectiv'. Există şi o noţiune limită de raţiune în sistemul lui Hegel; aceasta este stabilă, reprezintă stadiul final în care Spiritul devine pe deplin conştient de sine, care nu poate fi depăşită. Când relativiştii de astăzi îl 'naturalizează' pe Hegel prin conceptul-limită de raţionalitate adevărată, ei îi deturnează doctrina spre un fel de relativism cultural.

Nici un relativist nu vrea să fie relativist cu privire la *orice*. Cum ar putea gânditorii francezi pe care îi am în vedere să limiteze propriul lor relativism? Răspunsul diferă de la un gânditor la altul. În cazul unui marxist ca Althusser, răspunsul adoptat este o versiune a teoriei 'intereselor de clasă': toate 'ideologiile' sunt produsul unor factori iraţionali, dar acele ideologii care rezultă din interesele clasei muncitoare (de acum) sunt 'juste' şi tind în direcţia eliberării omului, în vreme ce ideologiile care împărtăşesc interesele claselor exploatatoare sunt 'nejuste' şi generează mizerie. Dar Althusser distinge între reprezentanţii anteriori şi cei actuali ai clasei muncitoare – punct de vedere relativist prin refuzul de a spune că chiar ideologia marxistă (ideologia 'clasei muncitoare') este *adevărată* sau *mai aproape de adevăr decât ideologia burgheză*. Ideologiile pot să fie 'juste' sau 'nejuste', conform lui Althusser, dar nu adevărate sau false[3]. ('Adevărat' sau

[3] După Althusser, 'Propoziţiile filosofice sunt teze'. 'Tezele filosofice pot fi susţinute *negativ* ca propoziţii dogmatice în măsura, în care ele nu

'fals' se aplică în cazul ştiinţelor de laborator şi în cazul acelor judecăţi pentru care există condiţii de testare empirică determinate). Foucault pare să încline, şi el, spre soluţia intereselor de clasă în ultimele lucrări, deşi este dificil să fim siguri. Nucleul unei asemenea doctrine, cel puţin în forma radicală a lui Althusser, este că tentativa de a salva afirmaţia relativismului radical că nici o ideologie nu poate fi raţională în vreme ce se acreditează ideea că unele ideologii (de preferat una singură – cea marxist-leninistă în cazul lui Althusser) pot fi bune, iar distincţia între ideologii 'bune şi 'rele' sau 'drepte' şi 'nedrepte' are alte temeiuri decât acceptabilitatea raţională. Deşi orice ideologie este adoptată din motive iraţionale sau non-raţionale, unele cauze neraţionale (interesele clasei muncitoare) sunt bune şi produc ideologii bune (prin definiţie?) în vreme ce alte cauze non-raţionale sunt rele şi dau naştere la ideologii rele. În loc de a judeca ideologiile potrivit *raţiunilor* (care sunt întotdeauna raţionalizări) ar trebui să le judecăm prin prisma cauzelor lor.

Această manieră de limitare a propriului relativism este clar nefuncţională, deoarece, care este judecata pe care se bazează teza că victoria 'intereselor clasei muncitoare' va avea consecinţe dezirabile cum ar fi o lume fără rasism şi război şi nu totalitarism şi imperialism deghizate în 'socialism'? Dacă răspunsul este că o asemenea variantă nu este socialism 'cu adevărat', nu reflectă cu adevărat

---

sunt susceptibile de demonstraţie într-o accepţiune *strict* ştiinţifică (aşa cum se discută în matematică şi logică), dar nici nu pot fi demonstrate în sens *strict* ştiinţific (precum în ştiinţele experimentale)... Tezele filosofice, odată ce nu pot fi nici demonstrate, nici dovedite, nu pot fi considerate "adevărate" (demonstrate sau dovedite ca în matematică sau fizică). Despre ele se poate spune numai că sunt "juste"', *Philosophie et philosophie spontanée des savants*, pp. 13–14, Maspero (1967).

interesele clasei muncitoare, atunci pe ce judecată se bazează teza că orice politică sau instituţie particulară (cum ar fi Partidul Comunist Francez) va promova interesele 'adevărate' ale clasei muncitoare şi 'adevăratul' socialism? Dacă asemenea convingeri pot fi justificate raţional, atunci nu orice ideologie este iraţională; dacă nu pot fi justificate raţional, atunci afirmaţia că vreo instituţie sau politică este 'justă' trebuie să rămână la fel de iraţională ca orice altă afirmaţie ideologică. Dacă oamenii ar considera problemele ideologice drept absurde, atunci cred la fel şi despre ideologiile care izvorăsc din 'interesele clasei muncitoare'.

Înapoindu-ne la Foucault şi lăsând deoparte semnalele din lucrările sale recente că a devenit mai radical, motivul pentru care selectează anumite cazuri este *tocmai* pentru a sugera tezele non-raţionale (de fapt, *iraţionale*) pe care le formulează asupra motivelor reale datorită cărora oamenii adoptă poziţii ideologice. De această dată, noţiunea de ideologic este foarte largă; nu numai comunismul, fascismul, democraţia, dreptul divin al regilor etc. sunt în discuţie. Convingerea că cineva este 'bolnav' şi are nevoie de 'tratament', convingerea că cineva este 'criminal' şi ar trebui, dacă este posibil, să fie 'reeducat', precum şi multe, multe alte convingeri din viaţa de fiecare zi sunt 'ideologice'. Într-adevăr, pentru ochiul de vultur al marxistului sociologizant sau al filosofului francez, aproape *orice* convingere este 'ideologică'. Poate că convingerea 'dacă scap acest pahar se va sparge' ar fi recunoscută ca neutră ideologic, dar asemenea cazuri sunt puţine.

S-ar putea crede că îmi scapă adevărata teză pe care o susţine Foucault. După cum mărturiseşte el însuşi, aceasta nu este că perspectivele ideologice din trecut erau nebuneşti sau iraţionale, ci mai degrabă că toate ideologiile, în sensul foarte larg în care foloseşte termenul, inclusiv

ideologiile noastre de astăzi, sunt relative la cultură. El încearcă să demonstreze că fiecare cultură trăieşte, gândeşte, face dragoste ghidată de o mulţime de presupoziţii cu determinare non-raţională. Dacă ideologiile precedente par acum 'iraţionale' este pentru că le judecăm prin prisma noţiunii noastre de raţionalitate.

Ceea ce este neclar în abordarea propusă de Foucault este că determinanţii pe care el şi alţi gânditori francezi iau în discuţie sunt *iraţionali în lumina situaţiei de astăzi*. Dacă ideologia actuală este produsul forţelor care sunt iraţionale prin prisma propriilor standarde, atunci este incoerentă. Gânditorii francezi nu sunt *numai* relativişti cultural; ei atacă noţiunea actuală de raţionalitate din interiorul ei, şi tocmai acesta este aspectul pe care cititorul îl poate găsi ciudat.

Relativismul cultural nu este o doctrină nouă. Antropologii ne-au predicat relativismul cultural de când există ei ca antropologi. Dar ar fi o eroare să asimilăm relativismul lui Foucault cu relativismul lor.

Atunci când un antropolog insistă asupra relativismului el aduce în discuţie practici şi credinţe ale vreunui trib exotic care ni se par iniţial respingătoare şi iraţionale şi încearcă să arate că rolul lor este de a asigura bunăstarea şi coeziunea grupului. Pe scurt, el demonstrează (cu urmarea că exemplul este rezonabil) că, deşi practicile avute în vedere sunt considerate greşite sau iraţionale în societatea noastră, sunt raţionale şi corecte în alte circumstanţe sociale.

Adesea, antropologii trag o concluzie eronată din propriile lor exemple (iar unele dintre acestea sunt mai puţin evidente decât cred antropologii). Ei vor spune 'totul este relativ', înţelegând că nu există fapte care să conteze ca fiind corecte sau greşite prin ele însele. Aşa cum mi-a sugerat Richard Boyd, aici se ascunde un motiv politic: să

ne convingă să stopăm distrugerea culturilor primitive prin punerea sub semnul întrebării a superiorităţii propriei noastre raţionalităţi şi morale. Din păcate, acest argument este foarte confuz. Exemplele antropologilor (atunci când sunt bune) arată că eronat şi corect, de pildă, sunt relative la circumstanţe, şi *nu* arată că nu ar exista corect sau eronat de loc, chiar în circumstanţele specificate. Argumentul împotriva imperialismului cultural ajunge să fie următorul: alte culturi nu sunt în mod obiectiv mai rele decât ale noastre (deoarece nu există asemenea lucruri precum mai binele sau mai răul obiectiv, după ei); deci ele sunt *la fel de bune cu a noastră*; prin urmare este greşit să le distrugem.

Argumentul este echivoc. Concluzia cere ca, *tot aşa de bune* să însemne *obiectiv* la fel de bune (cel puţin, potrivit standardelor noastre); dar din nonexistenţa valorilor obiective nu poate urma că orice este (în sensul cerut) 'tot atât de bun' ca altceva, ci mai degrabă că nu există ceva de felul 'la fel de bun'. Dacă valorile ar fi într-adevăr arbitrare, atunci nu ar trebui să distrugem culturile după bunul nostru plac?

Din fericire, există temeiuri mai bune decât negarea valorilor obiective pentru a critica imperialismul cultural. Scopul antropologilor este unul bun, dar este susţinut printr-un argument nepotrivit. Alt termen echivoc este cel care priveşte noţiunea de a fi relativ, pe care îl are în vedere Dewey prin 'relativism obiectiv'. Anumite lucruri sunt corecte – *corecte obiectiv* – în anumite circumstanţe şi greşite – *greşite obiectiv* – în alte circumstanţe, iar cultura şi mediul constituie circumstanţe relevante. Cu privire la acestea, antropologul are dreptate. Dar nu este acelaşi lucru cu relativitatea valorilor în sensul că sunt simple chestiuni de gust.

Mai mult, eliberat de confuzii conceptuale, argumentul antropologului trebuie acceptat. Observaţiile

sale ar trebui aprobate deoarece au rolul de a trezi sensibilităţile şi de a demonta prejudecata superiorităţii culturale. Dar comparaţia argumentării lui Foucault cu aceea a antropologului pune în evidenţă diferenţa dintre ele: Foucault nu argumentează că practicile trecute erau *mai* raţionale decât par a fi, ci susţine că toate practicile sunt mai puţin raţionale şi sunt, de fapt, determinate de puteri iraţionale şi egoiste. Asemănarea dintre această doctrină şi relativismul cultural al antropologului este doar superficială.

Poziţia pe care am discutat-o constituie o tentaţie intelectuală care este produsul creşterii cunoaşterii noastre privind mecanismul sociologic şi psihologic al sensibilităţii. Cunoaşterea şi sensibilitatea sunt parţial virtuale şi parţial reale; tentaţia este capcana de a concluziona că orice argument raţional este o simplă raţionalizare şi atunci încercăm să argumentăm *raţional* pentru o asemenea poziţie.

Dacă orice argument raţional ar fi o simplă raţionalizare, atunci nu numai că are sens să încercăm să argumentăm raţional pentru orice punct de vedere, ci are sens să *susţinem* orice punct de vedere. Dacă eu consider propria mea adeziune sau respingere ca un comportament anormal, atunci ar trebui să abandonez o asemenea atitudine – ceva pentru care nu există nici o adeziune sau respingere raţională, numai o parodie anormală de discurs raţional, nu poate fi numită *judecată*. Precum Sextus Empiricus, care trăgea concluzia că propriul scepticism nu poate fi exprimat printr-o judecată (deoarece chiar judecata 'Nu ştiu' nu poate fi *cunoscută*), relativistul modern, dacă ar fi consistent (dar cum ar putea cineva care consideră consistenţa un nonsens să susţină consistent o doctrină?) ar trebui să considere propriile afirmaţii numai ca expresie a unui sentiment.

De aici nu rezultă că nu am putem gândi raţional şi corect că *unele* dintre convingerile noastre sunt iraţionale, ci doar că există limite dincolo de care insistenţa că suntem neputincioşi intelectual devine neinteligibilă. O asemenea doctrină este cea lansată de Foucault; facem efortul de a fi imparţiali; încercăm să adoptăm ceea ce Popper numea 'o atitudine critică' şi să căutăm evidenţe şi argumente care ar putea să elucideze înţelesurile presupuse, chiar dacă acestea nu se potrivesc cu propria noastră opţiune. Nici o asemenea precauţie nu are sens dacă nu credem că aceste practici de discuţie şi comunicare şi aceste virtuţi ale criticii şi imparţialităţii servesc pentru a demonta convingerile iraţionale, dacă nu dintr-o dată, măcar gradual, în timp, şi să le înlocuiască prin convingeri care merită încrederea. S-ar putea ca raţionalitatea să nu fie reductibilă la un canon sau la o mulţime de principii, *dar* trebuie să existe un set de virtuţi cognitive care să ne ghideze.

S-ar putea obiecta, totuşi, că această poziţie nu este cu totul imparţială. Rudolf Carnap şi John Newman sunt ambii gânditori responsabili şi atenţi, şi acceptă asemenea virtuţi cognitive, dar nici unul nu crede că ar putea fi convins de celălalt dacă ar trăi în acelaşi loc şi s-ar întâlni. Faptul că nu există o cale pentru a soluţiona disputele încât toată lumea să fie satisfăcută nu demonstrează că nu există mai rău sau mai bine într-un asemenea caz. Mulţi cred că pentru Newman catolicismul era o obsesie; şi mai mulţi filosofi sunt de acord că, strălucitor cum era, Carnap a apelat la multe argumente slabe. Dacă facem asemenea judecăţi arată că folosim principii regulative privind corectitudinea, cât şi criterii comparative şi considerăm că există un standard asupra modului în care gânditorii răspund unui asemenea ideal. Unii vor spune 'ce-i cu asta? nu suntem mai buni dacă reuşim să rezolvăm o dispută decât dacă nu ar exista nici o noţiune de

acceptabilitate raţională exterioară punctelor de vedere în dezbatere la care să se poată apela!' Asta este adevărat atunci când suntem în faţa unei dispute irezolvabile, cum este cea imaginată între Newman şi Carnap; dar nu este adevărat că, dacă am abandona ideea că există asemenea lucruri precum imparţialitate, consistenţă şi responsabilitate, ar fi la fel de bine chiar dacă nu putem decât să le aproximăm în viaţa zilnică sau în practică şi ajungem să considerăm că există numai convingeri subiective şi nici un fapt despre care aceste 'simţăminte subiective' să fie corecte.

Poate că analogia pe care am trasat-o (ocazional) între disputa filosofică şi disputa politică poate fi de ajutor. Unul dintre colegii mei este un cunoscut apărător al punctului de vedere că orice guvern care cheltuie pentru bunăstare este moral neacceptabil. În optica sa, până şi sistemul public de învăţământ este eronat moral. Dacă sistemul public de învăţământ împreună cu legea educaţiei obligatorii (care, cred, ar privi-o tot ca un amestec brutal al guvernului asupra libertăţii individului) ar fi desfiinţate, atunci cele mai sărace familii nu şi-ar putea permite să-şi trimită copiii la şcoală şi s-ar ajunge la creşterea numărului de analfabeţi; dar situaţia poate fi rezolvată, din perspectiva lui, prin intermediul milei private. Dacă oamenii nu ar fi destul de miloşi pentru a preveni analfabetismul de masă (sau înfometarea în masă a oamenilor în vârstă) ar fi un lucru foarte rău care nu legitimează deciziile unui asemenea guvern.

Din punctul *meu* de vedere, premisele *sale* fundamentale – absolutizarea dreptului la proprietate, de pildă – sunt contraintuitive şi nu pot fi argumentate suficient. Din punctul *său* de vedere, *eu* sunt prizonierul unei filosofii paternaliste pe care el o consideră în neconcordanţă cu drepturile omului. Aici este o diferenţă profundă de vederi, dar mai degrabă în ce priveşte 'filosofia

politică' decât 'politica'. Dar multe disensiuni politice sunt create de disensiunile din filosofia politică, deşi nu întotdeauna atât de incompatibile ca aceasta.

Ce se întâmplă într-un asemenea caz? Când ambele variante sunt susţinute în mod inteligent este posibil uneori să se circumscrie sursa neînţelegerii. Adesea, când neînţelegerea este mai puţin fundamentală, ambele părţi şi-ar putea schimba poziţia mai mult sau mai puţin. Dacă nu se ajunge la înţelegere, părţile se pot mulţumi cu un compromis mai mult sau mai puţin acceptabil.

Asemenea discuţii inteligente între oameni cu vederi diferite sunt, din păcate, rareori întâlnite în zilele noastre; dar, cu atât mai bine atunci când se întâmplă. Iar cele două atitudini, într-o asemenea dispută, interacţionează într-un mod interesant. Pe de o parte, trebuie apreciate şi recunoscute virtuţi intelectuale de cea mai mare importanţă: minte deschisă, luarea în considerare a argumentelor şi motivaţiilor, capacitatea de a accepta critica etc. Dar care este situaţia cu fundamentele asupra cărora nu se poate cădea de acord? Am fi nesinceri dacă am susţine că aceste două poziţii nu pot fi comparate ca fiind mai bune sau mai rele. Nu este o simplă problemă de gust dacă cineva crede în obligaţia comunităţii să-i trateze pe membrii săi cu compasiune, dându-i întâietate faţă de dreptul la proprietate; la fel şi partenerul meu. Fiecare dintre noi îl consideră pe celălalt ca lipsit, la un anumit nivel, de sensibilitate şi percepţie. Pentru a fi sincer până la capăt, fiecare dintre noi se caracterizează printr-un fel de *respingere*, nu pentru inteligenţa celuilalt – fiecare îl consideră pe celălalt de cea mai bună calitate intelectuală – nici faţă de celălalt ca persoană – deoarece respect onestitatea, integritatea şi bunătatea celuilalt, chiar mai mult decât în cazul altor colegi care sunt de acord cu mine

în privinţa opiniilor politice – ci pentru complexul de emoţii şi judecăţi ale fiecăruia.

Dar nu am devenit mai puţin sincer de data asta? Spun că îi respect lui Bob Nozick inteligenţa, şi aşa este. Spun că îi respect caracterul, şi aşa se întâmplă. Dar când simt respingere pentru emoţiile şi judecăţile lui, oare nu este vorba de respingere pentru *el*?

Există un lucru dureros pe care politeţea mă opreşte să-l examinez cu părtinire în legătură cu atitudinile noastre de acceptare sau de respingere. Niciunuia dintre noi nu-i place şi nu respectă *orice* despre *oricine* (cel puţin în ceea ce ne priveşte). Nu există nici o contradicţie între a avea un respect deosebit pentru cineva şi, totuşi, a considera că are o slăbiciune intelectuală sau morală, aşa cum nu este nici o contradicţie între a te respecta pe tine dar a considera că ai o slăbiciune intelectuală sau morală (sau emoţională).

Vreau să spun că putem întâlni dispute între oponenţi de toate felurile: între un oponent care are virtuţi intelectuale precum minte deschisă, respect pentru raţiune şi dispoziţie pentru autocritică şi unul care nu le are; între un oponent care are un bagaj de cunoştinţe impresionant şi unul care nu are; între un oponent care se predă sentimentelor şi fanteziei (ceea ce fac oamenii obişnuiţi când discută despre politică) şi unul care raţionează cu atenţie. Iar atitudinea ambivalentă cu privire la respingere şi respect este una sinceră: respect pentru virtuţile intelectuale ale celuilalt; respingere pentru slăbiciunile lui emoţionale (conform cu standardele proprii, întotdeauna încep de la celălalt). 'Respingere şi respect' poate părea exagerat. Şi ar fi exagerat în cazul când respingerea l-ar viza pe celălalt ca persoană şi nu

numai pentru un complex de sentimente şi judecăţi. Dar este o atitudine mai sinceră decât *relativismul fals*, adică decât ipoteza că nu există nici o raţiune pentru ca individul să simtă că punctul său de vedere este rezonabil, pe când al celuilalt este iraţional.

Ar fi de ajutor să coborâm de la nivelul abstract la care am discutat şi să ne întoarcem din nou la exemple relativ simple. Să ne oprim la judecata pe care oamenii obişnuiţi sunt gata să o emită cel mai des, că pacea este preferabilă războiului. (Asemenea judecăţi nu sunt niciodată discutate de Foucault, aşa cum nu au fost descrise de Swift, pentru cel puţin un motiv: ambii sunt *autori satirici*. Numai nebunia socială îi interesează, nu şi – atunci când există – sănătatea socială). Nu există îndoială privind motivaţiile unei asemenea judecăţi. Suntem prea familiari cu ororile războiului, cu ceea ce se întâmplă în război cu adulţii şi copiii, cu combatanţii şi necombatanţii, cu pământul şi vegetaţia. Chiar dacă o asemenea judecată izvorăşte parţial din interesul propriu, asta *nu* o face iraţională; ci dimpotrivă.

Chiar dacă o întreagă populaţie ar judeca altfel, că războiul este preferabil păcii, nu s-ar găsi raţiuni care să motiveze o asemenea atitudine. Agresiunea şi fantezia pot împinge oamenii spre vărsare de sânge. Dar, iarăşi, de aici nu rezultă că toate judecăţile de valoare sunt iraţionale, ci numai unele; şi este foarte greu să spunem care şi când merită încrederea noastră (iată de ce susţin importanţa imparţialităţii şi spiritului critic între virtuţile cognitive).

Faptul că unele judecăţi de valoare sunt raţionale şi obiective nu înseamnă că discuţiile noastre despre capitalism, democraţie, socialism, drepturi, autonomie etc., nu sunt de multe ori fără sens. Chiar atunci când ceea ce vrem să spunem despre ceva general este corect, avem uneori probleme în a ne exprima îndeosebi când nu suntem obişnuiţi cu ideile abstracte. Aici este inclus cazul antropologului care, atunci când spune că nu există valori obiective, vrea să spună că valorile sunt relative la circumstanţele cazului respectiv. Chiar dacă avem succes în a exprima ce dorim să spunem, există forţe iraţionale care tind să guverneze judecata noastră. Nu neg că o asemenea putere poate corupe judecata; dar resping că suntem neajutoraţi în faţa forţelor iraţionale, încât ar fi o dovadă de delăsare să nu încercăm să judecăm cu inteligenţă şi dreptate. A spune că *putem* fi raţionali este altceva decât a spune că suntem infailibili. Dimpotrivă, aşa cum subliniază Iris Murdoch, năzuinţa pentru o instanţă raţională este esenţialmente ceva progresiv şi 'perfectibil la infinit'[4].

Afirmaţiile de până acum ar putea fi acceptate de un adept inteligent al relativismului moral. Nu este nevoie ca un relativist să fie preocupat de nedeterminarea raţionalităţii tuturor judecăţilor de valoare, sau de apărarea tabloului istoriei zugrăvit de Foucault, ca o serie discontinuă de discursuri şi de ideologii care se impun fără vreun motiv raţional. Un relativist mai modest s-ar mulţumi cu teza lui Dewey că unele valori sunt relative

[4] *The Sovereignity of Good*, p. 23.

obiectiv – în circumstanţe date, naturale sau istorice, faţă de cei care le evaluează. Un asemenea relativist modest menţine teza relativităţii valorilor. 'Obiectivitatea' pe care o neagă nu este obiectivitatea pe care o afirma Dewey[5], că orice judecată este întemeiată faţă de domeniul său, fiind mai degrabă obiectivitatea pe care o afirmă Platon, obiectivitatea nelegitimă (cum ar spune relativiştii) care priveşte discursul despre un reper absolut, deosebit de orice circumstanţă şi valid în orice circumstanţă.

Dacă nu mă mulţumesc cu acest relativism modest, dacă nu mă simt atras de etica lui Dewey, nu se întâmplă pentru că tânjesc după Absolut. Când afirm că crima şi suferinţa oamenilor nevinovaţi sunt greşite, nu îmi pasă dacă această judecată ar fi validă pentru orice constituţie şi psihologie imaginabilă. Dacă există fiinţe, să spunem pe Alpha Centauri, care nu simt durere şi cărora nu le pasă de moartea unui individ, atunci agitaţia noastră despre 'crimă şi suferinţă' le-ar părea prea mult zgomot pentru nimic. Dar asemenea fiinţe nu pot înţelege comandamentele morale prezente aici. Dacă 'obiectivitatea' noastră este obiectivitatea umană, comandamentele respective sunt suficient de obiective.

Doctrina lui Dewey despre 'relativitatea obiectivă' nu poate da seama de cazul nazismului (deşi Dewey nu ar fi ocolit subiectul). Vreau să spun că ţelurile naziştilor erau cu totul eronate, iar afirmaţia că 'aceasta este adevărat relativ la interesele noastre şi fals relativ la interesele naziştilor' este tocmai tipul de relativism moral

[5] Vezi teoria valorii a lui Dewey din *The Enciclopedia of Unified Sciences*, vol. II, nr. 4, Universitatea Chicago (1939).

pe care doresc să-l resping. Relativismul obiectiv pare a fi doctrina *corectă* pentru multe situaţii; dar nu pentru cazurile în care drepturile şi datoriile se manifestă atât de drastic încât alegerea trebuie făcută între corect şi eronat, între bine şi rău.

Într-adevăr, există teza că noţiunea modernă instrumentală de raţionalitate este ea însăşi 'relativist obiectivă'. Nucleul acestei noţiuni este o simplă dihotomie: alegerea scopurilor nu este *nici raţională nici iraţională* (deşi este nevoie de un nivel minim de consistenţă), în vreme ce alegerea mijloacelor este *raţională în măsura în care este eficientă.* Raţionalitatea este un predicat pentru mijloace, nu pentru finalităţi şi este în totalitate unit cu eficienţa. Astfel, preferinţa lui Jones pentru vanilie faţă de ciocolată într-o situaţie particulară ar fi raţională pentru Jones dacă îi generează satisfacţia urmărită. Această concepţie, care ne întoarce la dictumul lui Hume: 'raţiunea este şi trebuie să fie sclava pasiunilor', care l-a influenţat puternic pe Bentham, este larg recunoscută de cercetătorii moderni în ştiinţele sociale. Ea joacă un rol important în teoria bunăstării economice şi în multe alte domenii. Noţiunea economiştilor moderni *Pareto optimum* este o încercare de optimizare economică unde contează numai eficienţa mijloacelor şi nu presupune judecăţi de valoare privind scopurile agenţilor economici; această noţiune este astăzi de interes deoarece presupune că alegerea mijloacelor este subiect pentru critica raţională, în timp ce alegerea scopurilor nu este.

Această concepţie pierde mult din atractivitate când vedem că teorii psihologice simple nu i se supun. În

doctrina benthamistă, scopurile, finalităţile, ţelurile sunt considerate fie ca parametri individuali constanţi (învăţarea este descrisă ca un proces cu atât mai reuşit cu cât ajunge să estimeze consecinţele acţiunilor şi să conducă la atingerea cât mai eficientă a scopurilor, dar nu ca un proces de dobândire a noi scopuri) sau prin parametri individuali care, chiar dacă nu sunt constanţi, se schimbă numai ca rezultat al factorilor neraţionali şi pe care teoria nu îi poate surprinde.

Bernard Williams[6] a subliniat că există un număr de modalităţi în care scopurile individuale, şi nu numai mijloacele alese să le atingă, pot fi supuse criticii raţionale; căi care devin vizibile în momentul în care depăşim domeniul psihologiei lui Bentham.

Teoria lui Bentham lasă loc situaţiei în care un individ poate fi convins să abandoneze un scop (sau oricum, să abandoneze urmărirea lui) prin critică raţională: când a greşit estimarea consecinţelor aşa încât *costurile* atingerii scopului sunt mai mari decât cele anticipate. Aceasta deschide calea spre o problemă care are de-a face tot atât de mult cu imaginaţia ca şi cu inteligenţa propoziţională: ce s-ar întâmpla în planul experienţei cu atingerea scopului? Mulţi oameni care urmăresc un scop, odată ce l-au atins nu simt plăcerea pe care o aşteptau. Chiar în interiorul schemei lui Bentham ar fi posibil ca decizia raţională să fie abordată luând în considerare posibilitatea unei estimări eronate a *plăcerii* resimţite în legătură cu atingerea unui scop. În acest fel, nu numai mijloacele, dar şi scopurile ajung să fie criticate raţional.

---

[6] Aici expun pe scurt o conferinţă intitulată 'Internal and External Reasons', susţinută la Harvard, în noiembrie 1978.

Uneori, oamenii *adoptă* scopuri pe care nu le pot urmări mental. Sau, chiar dacă se gândesc la ele (sau cineva le sugerează) le poate lipsi imaginaţia (iar imaginaţia!) să vizualizeze ce pot obţine dacă aceste scopuri ar fi aşa cum par – cu atât mai mult, dacă aceste scopuri sunt trăsături de caracter, cum ar fi dezvoltarea gustului pentru poezie. Cel care preferă pushpin poeziei poate nu este în stare să-şi imagineze ce ar simţi dacă ar fi receptiv la adevărata poezie şi dacă şi-ar da seama că inteligenţa sa ar putea creşte sau imaginaţia s-ar putea îmbogăţi, poate şi-ar da seama că a făcut o *greşeală*.

Este important că abilitatea de a critica raţional scopurile cuiva poate depinde tocmai de bogăţia imaginaţiei tot atât de mult ca de abilitatea de a deosebi judecăţile adevărate de cele false. De asemenea, are importanţă faptul că scopul analizat poate fi o trăsătură de caracter, nu numai un lucru sau un eveniment.

Există şi alte căi, pe lângă estimarea greşită a satisfacţiei unui scop sau a scopurilor alternative, prin care cineva poate alege greşit un scop. Williams a subliniat (invocând o observaţie care provine de la Aristotel) că adesea scopul este *general* (de pildă, 'să mă distrez deseară'), iar problema nu constă atât de mult în găsirea *mijloacelor* pentru a-l duce la îndeplinire, ci a găsi tipul de activitate care să constituie o *specificare* acceptabilă a acelui scop (cum ar fi 'să merg la cinematograf' sau 'să stau acasă să citesc o carte' etc.). Dacă cineva poate imagina specificaţii noi şi creative ale scopului sau numai specificaţii banale depinde mai ales de imaginaţie.

Problema, aşa cum a pus-o Williams, este că, chiar dacă înlocuim psihologia îngustă a lui Bentham cu o

abordare corectă a tuturor acestor lucruri, rămâne loc pentru relativism. Exemplul lui Williams era situaţia ipotetică a unui tânăr al cărui tată doreşte să-l determine să aleagă o carieră militară. El invocă tradiţia familiei (bărbaţii au fost ofiţeri de multe generaţii) şi patriotismul, dar degeaba. Chiar când tânărul se autosugestionează că ar putea să-i placă să fie ofiţer, nu găseşte nimic în acest scop care să-l atragă. Nu este scopul *său*, dar nu pentru că nu ar avea suficientă inteligenţă sau imaginaţie.

Cazul naziştilor ar putea fi asemănător. Să presupunem că naziştii au câştigat războiul, astfel că nu se poate invoca înfrângerea Germaniei ca motivaţie practică să nu fim nazişti. Poate unii nazişti nu cunosc consecinţele nazismului, suferinţele pe care le-a adus etc. Poate că unii dintre ei nu ar fi fost nazişti dacă ar fi avut suficientă inteligenţă şi imaginaţie încât să aprecieze consecinţele sau să aprecieze alternativa unei vieţi mai bune. Dar fără îndoială, mulţi ar fi rămas nazişti deoarece nu le pasă de suferinţele pe care le generează acţiunile lor şi deoarece, pentru ei, nu prezintă importanţă că ar fi avut o viaţă mai bună în alt mod, nu ar fi procedat cum a făcut tânărul din povestirea lui Williams. Nu există finalitate *în ei* la care să apeleze, nici actuală, nici potenţială, pe care ar putea-o alege dacă ar fi mai imaginativi sau mai inteligenţi. Chiar ignorând 'psihologia' lui Bentham suntem confruntaţi cu problema relativismului moral.

Să discutăm şi un caz mai puţin important decât al naziştilor. Să ne imaginăm o societate de fermieri care, dintr-un motiv oarecare, au un dezinteres total faţă de artă, faţă de ştiinţă (cu excepţia realizărilor care îi pot

ajuta în munca de la fermă), faţă de religie, pe scurt, faţă de orice domeniu spiritual sau cultural. (Nu vreau să sugerez că fermierii actuali ar fi precum cei imaginaţi aici). Nu trebuie să ni-i imaginăm pe aceşti oameni ca răi; ne imaginăm că ei conlucrează paşnic, sunt suficient de atenţi unul faţă de altul. Cititorul să-şi închipuie numai că interesele lor sunt limitate la scopuri minime, precum să aibă suficiente alimente, să aibă un adăpost cald şi alte plăceri simple cum ar fi petrecerea împreună a unor evenimente deosebite. Pe scurt, ne imaginăm că aceştia duc o existenţă apropiată de cea a animalelor şi nu îşi doresc să aibă un alt fel de existenţă.

Asemenea oameni nu sunt *imorali*. Nu există nimic nepermisibil în modul lor de viaţă. Dar tendinţa noastră naturală (numai dacă nu suntem sub stăpânirea relativismului radical) este să considerăm că stilul lor de viaţă este, pe undeva, de compătimit. Lor le lipseşte cu desăvârşire ceea ce Aristotel numea 'nobilitate'. Ei duc o viaţă de porci – porci prietenoşi, poate, dar totuşi porci, iar viaţa unui porc nu este o viaţă pentru un om.

În acelaşi timp – şi aceasta este problema – nu suntem tentaţi să spunem că aceşti oameni-porci ar fi în vreun fel *iraţionali*. Aceasta poate fi rezultatul lipsei noastre îndelungate de cultură privind modul în care benthamiştii folosesc termenii 'raţional' şi 'iraţional', dar s-ar putea să fie şi rezultatul dispoziţiei noastre actuale. Vieţile oamenilor-porci nu sunt atât de bune cât ar putea să fie, vrem să spunem, dar ele nu sunt *iraţionale*.

Deşi nu susţin că este doar o problemă de *gust* să avem o viaţă mai bună sau mai rea, nu văd cum se poate argumenta că este *raţional* să alegem viaţa mai bună şi că

este *irațional* să o alegem pe cea mai rea. Nu vreau să spun că asemenea alegeri conduc la urmarea că 'totul este *relativ*'; de data aceasta, întemeierea este dificil de descoperit.

Poate unele corecții la psihologia benthamistă sugerate de Bernard Williams ne-ar putea ajuta în *acest* caz. Să presupunem că oamenii-porci s-au născut cu un *potențial* uman normal (dacă nu, atunci viețile lor *nu sunt* 'mai rele decât ar putea fi' așa că nu suntem justificați să-i compătimim, ci numai, cel mult, să simțim puțină milă față de ei). În acest caz ei ar putea fi determinați să aprecieze aspectele artistice, științifice sau spirituale ale vieții; să trăiască o viață mai umană. Dacă vreunul dintre ei ar face astfel, fără îndoială că și alții ar prefera viața acestuia (deși le-ar părea mai puțin distractivă). Oameni care trăiesc vieți animalice se vor simți *rușinați* când vor ajunge la o viață mai umană; cei care trăiesc vieți mai umane nu ar avea sentimentul de rușine când ar ajunge într-o viață decăzută. Aceste observații sunt temeiuri pentru a crede că oamenii-porci se află într-un fel de eroare, într-un fel de sincopă *cognitivă*, potrivit lui Williams; temeiuri pentru a crede că ei nu se gândesc la scopuri alternative și temeiuri pentru a crede că ei nu au făcut niciodată compromisuri față de ei înșiși pentru a-și imagina cum s-ar simți în situația scopurilor alternative. Pe scurt, ei nu ar putea recunoaște că au *ales* o viață mai rea, deoarece nu au avut niciodată o concepție adecvată despre ce este mai bine.

Observațiile anterioare ar putea să ne edifice asupra noțiunii că viața *acestor* oameni este supusă criticii raționale, dar nu este la fel de clar cum trebuie aplicată,

faţă de situaţia naziştilor. (S-ar putea invoca tautologia că oricine nu *alege* o viaţă mai bună nu are o concepţie adecvată asupra ei; dar o asemenea manevră nu ne ajută prea mult.) Chiar în cazul oamenilor-porci, dacă printre ei ar fi oameni-porci *fanatici*, opunându-se oamenilor-porci obişnuiţi, atunci observaţia privind ruşinea nu ar avea loc. Fanaticii nu concep alternative la care s-ar putea raporta.

Reţinerea de a-i acuza pe oamenii-porci de un defect *raţional* (cu excepţia cazului când, cel puţin latent, nu reuşesc să atingă unele dintre scopurile *lor*) este urmarea recentelor critici ale noţiunii de raţional în cultura noastră, după cum este uşor să se stabilească din faptul că nici un filosof antic sau medieval nu ar vedea nimic straniu în a spune că dacă *A* este o viaţă mai bună decât *B*, atunci este *raţional*, cea mai bună raţiune posibilă, să se *aleagă A* mai curând decât *B*. Am pierdut abilitatea de a vedea cum bunătatea unui scop conduce la *raţionalitatea* alegerii lui.

Desigur, aceasta este o explicaţie extinsă a faptului că nu considerăm 'bunătatea' ceva obiectiv. În acest caz, suntem confruntaţi cu un cerc sau mai degrabă cu două curbe. Există cercul modern: concepţia instrumentalistă despre raţionalitate susţine afirmaţia că alegerea unei finalităţi bune, sau care nu este rea în mod absolut, nu conduce la iraţionalitate dacă acea finalitate nu este aleasă, ia această teză, la rândul ei, susţine afirmaţia că bunătatea sau răutatea nu sunt obiective, care, la rândul ei, susţine afirmaţia că noţiunea instrumentalistă de raţionalitate este singura inteligibilă. Există şi un arc tradiţional: raţiunea alege finalităţile pe temeiul *bunătăţii*

lor (în opoziţie cu 'pasiunea' care ne dictează finalităţi pe baza apetitului sau a 'înclinaţiei'); se ajunge la punctul de vedere că este *raţional* să se aleagă binele, care la rândul său întemeiază afirmaţia că bunătatea şi răutatea sunt obiective. De bună seamă, nu putem să ne întoarcem pur şi simplu la perspectiva medievală, oricât de conservatori am fi; dar, oare, cercul benthamist să fie o alternativă viabilă?

Impactul ştiinţei asupra concepţiei moderne privind raţionalitatea

Discuţia pe care am prezentat-o pare fără sfârşit – având deja mai multe decade – datorită, poate, presupunerii că raţionalitatea are un fel de prioritate asupra binelui. Problema se reduce la întrebarea dacă există vreun sens în care alegerea unei finalităţi rele poate fi numită iraţională, ca şi dacă bunătatea ar fi o dovadă de raţionalitate a judecăţii, dacă ar fi aleasă. A accepta o asemenea teză, îndeosebi când presupunerile despre raţionalitate sunt o colecţie largă de mituri culturale, înseamnă a adopta o prejudecată. În cele ce urmează propun inversarea termenilor şi să ne întrebăm nu atât cum este bunătatea raţională, ci dacă a fi raţional înseamnă a fi bun? Întrebarea privind valoarea raţionalităţii însăşi urmăreşte clarificarea naturii raţionalităţii, explici-

tarea presupunerilor pe care le acceptăm cu privire la raţionalitate, cu finalitatea de a putea stabili ce este eronat în modul în care am abordat până acum problema raţionalităţii

Să ne reamintim argumentul împotriva obiectivităţii judecăţilor de valoare a lui Max Weber, care a introdus distincţia modernă între fapt şi valoare, respectiv că nu se poate stabili adevărul unei judecăţi de valoare aşa încât să satisfacă toate *persoanele raţionale posibile*[1]. Chiar de la început se pune accentul pe imposibilitatea, sau pretinsa imposibilitate a judecăţilor de valoare. Raţionalitatea a supus valorile unei examinări continue multă vreme. În acest context, prin raţionalitate trebuie să se înţeleagă raţionalitatea ştiinţifică[2]; ea este rezultatul 'ştiinţei pozitive' despre

---

[1] Cf. îndeosebi 'Die Objektivität sozialwissenchaftlicher Erkenntnis', *Archiv für Sozialwiss. und Sozialpolitik*, vol. 19 (1904), pp. 24–87, şi 'Der Sinn der soziologischen und ökonomischen Wiessenschaften', *Logos*, VII (1917), pp. 49–88 şi 'Wiessenschaft als Beruf', Vortrag, 1919. Aceste trei texte au fost retipărite în *The Methodology of the Social Sciences*, Glencoe, Illinois, 1949.

[2] K.O. Apel îl înţelege pe Weber în acelaşi mod în 'The Common Presuppositions of Hermeneutics and Ethics: Types of Rationality Beyond Science and Technology', *Research in Phenomenology*, Nr. IX, 1979. Astfel, Apel scrie: 'Max Weber, cu toate acestea, propune şi un răspuns strict negativ cu privire la întrebarea mea despre tipurile posibile de raţionalitate dincolo de ştiinţa liberă de valori, şi de tehnologie. Acest răspuns a devenit paradigmatic, cred eu, pentru sistemul ideologic vestic. Weber restrânge domeniul înţelegerii metodologice la înţelegerea 'liberă de valori' care este centrată în jurul 'tipului ideal' al 'înţelegerii raţionale îndreptate spre scop' sau 'acţiunilor raţionale îndreptate spre scop'. Acum, asemenea acţiuni pot fi numite 'instrumentale'; în acele cazuri în care ele au succes, ele pot fi analizate sau reconstruite ca fiind bazate pe transpoziţiile de succes de la regulile dacă-atunci ale ştiinţei nomologice, în regulile dacă-atunci ale prescripţiilor tehnologice. În acest fel, Max Weber restrânge problema înţelegerii metodologice la tentativa de a surprinde *raţionalitatea mijloace-scopuri de natură tehnologică* (liberă de valori), dincolo de acţiunea umană. Aceasta este teza *raţionalităţii instrumentale* care constituie paradigma raţionalităţii în sens restrâns. Trebuie subliniat, totuşi, că pentru o *înţelegere*

care se spune că poate satisface toate persoanele raţionale. Un motiv pentru a valoriza raţionalitatea este clar. Raţionalitatea ştiinţifică ajută, fără îndoială, la atingerea scopurilor. În vreme ce puţini oameni educaţi ar subscrie la punctul de vedere că trebuie să urmăm ştiinţa numai de dragul succesului tehnologic, nu este nici o îndoială că succesul tehnologic al ştiinţei este covârşitor în cel mai literal sens al cuvântului. Trăim o continuă serie de revoluţii tehnologice care ne reamintesc mereu de forţa ştiinţei de a modela vieţile noastre – 'revoluţia industrială', 'revoluţia electronică'. Chiar înainte de revoluţia industrială, succesul fizicii newtoniene a impresionat numeroase minţi. De exemplu, când a început discuţia privind noţiunea de 'progres' în secolul al XVII-lea, adepţii progresului au întemeiat crezul lor pe afirmaţia că 'Newton l-a depăşit pe Aristotel'. Nimeni nu ar putea argumenta convingător că Shakespeare este un dramaturg mai bun decât vreunul dintre autorii de tragedii antici sau un poet ar mai bun decât Homer; dar pare imposibil de negat că ştiinţa lui Newton reprezintă un progres în cunoaştere în raport cu ştiinţa lui Aristotel.

Deşi enciclopediştii şi alţii s-au grăbit să generalizeze noţiunea de progres de la ştiinţă la instituţiile politice şi

---

raţională îndreptată spre scop în sociologie nu este necesar ca agentul să fie sigur de succesul transpunerii regulilor nomologice în maxime tehnologice cu privire la relaţii mijloace-scopuri. Pentru a înţelege acţiunile sale în lumina acestui tip de raţionalitate instrumentală, este suficient să ne asigurăm că era raţional ca agentul să acţioneze aşa cum a făcut-o sub anumite presupoziţii despre scopurile *sale* şi convingeri despre mijloacele *sale* sau căile ori strategiile ce trebuie urmate pentru atingerea scopurilor. În acest fel, devine o problemă empiric-hermeneutică privitoare la înţelegere să se găsească şi să se verifice în plan ipotetic acele relaţii intenţii-scopuri şi convingeri-mijloace din perspectiva agentului, în lumina cărora acţiunile sale pot fi înţelese ca raţionale în sensul raţionalităţii tehnologice mijloace-scopuri.

moralitate, generalizarea a părut pe atât de neclară în secolul al XX-lea pe cât de limpede era în secolul al XIX-lea.

Auguste Comte a construit o filosofie, pozitivismul, care celebrează succesul ştiinţei. Istoria, potrivit lui Comte, este povestea succesului: începe cu miturile primitive, acestea sunt purificate şi rafinate conducând spre religia înaltă, din care îşi trag seva sistemele metafizice ale lui Platon sau Kant, iar din metafizica de astăzi se ajunge la 'ştiinţa pozitivă'. De bună seamă că eroul acestei povestiri este ştiinţa. Ştiinţa atrage atenţia prin succesul ei intelectual şi tehnologic. Suntem impresionaţi îndeosebi de impactul pe care îl are asupra propriei noastre vieţi.

Rezultă un motiv pentru îndoiala că judecăţile de valoare ar avea un statut cognitiv, respectiv că nu pot fi verificate 'prin metode ştiinţifice' cum s-a spus mereu. De asemenea, trebuie menţionat faptul, care joacă un rol central în viziunea lui Foucault, că nu putem atinge acordul universal, sau măcar majoritar, în ceea ce priveşte chestiunile etice. Dacă avortul este corect sau greşit sau dacă homosexualitatea este corectă sau greşită nu par a fi chestiuni asupra cărora unele răspunsuri să întrunească acordul tuturor; în acelaşi timp, este larg răspândită convingerea că pentru o teorie ştiinţifică se poate demonstra corectitudinea, spre satisfacţia tuturor. Raţionalitatea ştiinţei este identificată uneori, cel puţin în parte, cu faptul că măcar predicţiile sale pot fi dovedite public; că oricine poate fi de acord că aceste rezultate sunt obţinute şi fenomenele prezise de teorie au loc. De bună seamă că aici apare un cerc: dacă *identificăm* procedurile raţionale cu cele care conduc la concluzii la care aderă majoritatea, atunci argumentul lui Weber, chiar dacă este corect, că în chestiunile etice nu se poate ajunge la consens de către

oamenii raţionali, ar însemna că nu se poate ajunge la consens între cei care folosesc metodele care sunt prevăzute că ar conduce la asentimentul majorităţii sau al unei majorităţi covârşitoare. Adică, modul în care determinăm că judecăţile de valoare nu pot satisface toţi oamenii *raţionali* nu este împărtăşit de majoritatea acestor oameni. Prin urmare, nu se obţine un test pentru raţionalitate; formularea lui Weber sugerează că, într-un fel, mai întâi socotim care sunt membrii raţionali ai unei populaţii şi apoi vedem dacă ei sunt sau nu de acord asupra adevărului unei judecăţi de valoare. Dar în realitate niciodată nu are loc un asemenea recensământ. Exemplele lui Weber (cu mandarinii chinezi şi altele) arată numai că judecăţile de valoare nu pot fi verificate în aşa fel încât să satisfacă toţi oamenii educaţi sau inteligenţi (care sunt întotdeauna aceiaşi cu oamenii *raţionali*). Într-o formă deghizată, argumentul lui Weber este argumentul majorităţii; el apelează la faptul că oamenii educaţi pot cădea de acord asupra 'ştiinţei pozitive', în vreme ce un asemenea argument nu poate fi adoptat în cazul valorilor etice. Este interesant contrastul acestei teze cu cea a lui Aristotel care acceptă că în etică ar trebui să încercăm obţinerea acordului celor mulţi, dar trebuie să fim conştienţi că adesea este imposibil. Uneori suntem capabili să-i convingem numai pe cei înţelepţi; şi de bună seamă, trebuie să avem încredere în recunoaşterea socială a celor înţelepţi.

De multe ori nu se obţine un acord covârşitor asupra adevărului unei teorii ştiinţifice. Cei mai mulţi oameni sunt ignoranţi în ce priveşte ştiinţa şi multe teorii, îndeosebi din ştiinţele exacte unde este cerută prea multă matematică, rămân neînţelese. În acelaşi timp, cei mai mulţi oameni nu sunt afectaţi de această situaţie. Deoarece, conform operaţionalismului slab, care a devenit filosofia de

lucru a multor oameni de ştiinţă, conţinutul teoriei ştiinţifice constă în consecinţe testabile, iar acestea pot fi exprimate în judecăţi de forma *dacă efectuăm cutare şi cutare acţiuni, atunci ajungem la cutare sau cutare rezultat observabil*. Judecăţile de această formă, dacă sunt adevărate, pot fi demonstrate prin repetarea experimentului potrivit, oricât de des. Este adevărat că această teză se confruntă cu o mare dificultate: experimentele sunt mai dificil de proiectat, realizat şi evaluat decât se crede, de către un om obişnuit. Dar nu există îndoială că, în acest fel, s-ar putea obţine un acord larg asupra adecvării experimentale a unor teorii din ştiinţele exacte. Acceptarea de către un neprofesionist a acestor teorii poate fi diferită de acceptarea experţilor, dar cel puţin experţii par a fi de acord.

La nivel intelectual, instrumentalismul nu reprezintă o teză mulţumitoare cu privire la raţionalitate. Rezultatele ştiinţifice au o valoare practică enormă dar, aşa cum am văzut deja, nici o persoană educată nu crede că ştiinţa este valoroasă numai de dragul aplicaţiilor practice. Chiar dacă ar fi aşa, de ce ar trebui ca *raţionalitatea* să fie redusă la aplicaţiile practice? Desigur că este util să avem un instrument care ne ajută să selectăm eficient mijloacele prin care atingem scopurile; dar este de asemenea valoros să ştim spre ce scopuri trebuie să ne îndreptăm. De aceea, nu este surprinzător că adevărul judecăţilor de valoare nu poate fi 'demonstrat raţional' dacă 'verificarea raţională' este limitată prin definiţie la stabilirea conexiunilor mijloace-scopuri. Dar de ce să rămânem la o concepţie atât de îngustă despre raţionalitate?[3]

---

[3] Atribuindu-i tocmai o asemenea concepţie îngustă despre raţionalitate lui Weber, Apel scrie (*loc. cit.*, p.73): 'Această preocupare a metodologiei lui Weber privind 'înţelegerea' este în concordanţă (mai

Apelul la majoritate nu este suficient. Pentru a ajunge la certitudine ar trebui dobândit acordul asupra a ceea ce trebuie stabilit ca adevărat. De asemenea, ar fi bine să se evite conflictele între susţinători. Dar oamenii admit că uneori este bine să ai încredere în judecata cuiva chiar dacă diferă de judecata majorităţii. Mulţi au avut succes încrezându-se în propria lor judecată, deşi era diferită de cea a majorităţii. Astfel, pentru unele subiecte, între care se află şi cele etice, condiţiile care trebuie îndeplinite pentru a fi evaluate sunt atât de complexe şi atât de imprecise că nu putem spera să ne bazăm pe ceva asemănător demonstraţiei sau definiţiei ştiinţifice, dar trebuie să avem încredere în percepţie şi în judecată. Una dintre cele mai înalte manifestări ale raţionalităţii ar trebui să fie abilitatea de a judeca în mod corect tocmai în acele cazuri în care nu avem speranţa unei demonstraţii care să satisfacă

---

mult sau mai puţin implicit) cu filosofia sa a istoriei. Deoarece chiar în contextul reconstrucţiei istoriei civilizaţiei occidentale, el porneşte de la ipoteza că măcar această parte a istoriei ar putea fi concepută ca un progres continuu în 'raţionalizare' şi, în acelaşi timp, ca un proces de depăşire a iluziilor sau, cum îi place să spună, 'dezvrăjire'. Prin 'raţionalizare' el înţelegea progresul în introducerea în toate sectoarele sistemului socio-cultural a raţionalităţii mijloace-scopuri, îndeosebi în sfera economicului şi a administraţiei birocratice, sub influenţa constantă a progresului în ştiinţă şi tehnologie. Prin procesul de 'dezvrăjire', pe de altă parte, Weber înţelegea, între altele, disoluţia tabelei de valori acceptate din perspectivă religioasă sau filosofică. El era pregătit să tragă consecinţe practice din această evoluţie pentru propriul său tablou asupra lumii în măsura în care el sugera că un gânditor riguros şi sincer ar trebui să accepte următoarele: progresul uman în sensul 'raţionalizării' îşi găseşte contraponderea în părăsirea ideii privind o cercetare raţională a valorilor sau normelor fundamentale în favoarea deciziilor preraţionale ale conştiinţei în faţa unui pluralism sau, cum spune Weber, 'politeism' al normelor sau valorilor.

majoritatea. Pare straniu ca, din imposibilitatea demonstrării, să rezulte iraţionalitatea convingerii[4].

Cu toată slăbiciunea intelectuală a acestor concepţii, atât instrumentalismul, cât şi ipoteza majorităţii sunt preferate de gândirea contemporană care este atrasă de succesul demonstrabil şi simte disconfort în faţa noţiunii de judecată şi înţelepciune. Nefiind sociolog, nu voi încerca să explorez de ce societatea industrială, atât în varianta capitalistă cât şi în cea socialistă, este sedusă de temele instrumentalismului şi ale majorităţii. Fără îndoială, faptul sociologic are de-a face cu echivalarea raţionalităţii cu raţionalitatea ştiinţifică, iar concepţia despre raţionalitate ştiinţifică se bazează pe demonstrarea legăturilor instrumentale cu satisfacerea (potenţială) unei majorităţi covârşitoare.

Deşi această viziune asupra raţionalităţii, în care concepţia despre raţionalitate ca ansamblu de metode (a căror natură rămâne de obicei vagă) care, oricare le-ar fi natura, conduc la descoperirea de conexiuni efective mijloace/scopuri şi recunoaşterea lor 'publică', nu se susţine intelectual, nu au lipsit încercările filosofice de a o face respectabilă. Una dintre aceste tentative dezvoltată pornind de la empirismul lui Locke, Berkeley şi Hume, este doctrina empiristă a lui Mill, numită de istoricii filosofiei, fenomenalism: în realitate există doar *senzaţii*. Obiecte obişnuite, cum sunt mesele şi scaunele, sunt numai mulţimi de regularităţi obiective în senzaţiile actuale şi posibile ale oamenilor. Cum se exprimă Mill, obiectele fizice sunt 'posibilităţi permanente de senzaţii'. O altă cale de a exprima aceeaşi idee constă în afirmaţia că discursul

[4] Într-adevăr, am văzut în capitolul 5 că teoria consensului pentru acceptabilitatea raţională se respinge singură!

despre ceea ce pare a fi lumea fizică, este de fapt, un discurs indirect despre senzaţii.

Meritul acestui punct de vedere, în viziunea susţinătorilor săi, era că le permite să spună cu claritate care este conţinutul, nu numai al ştiinţei, ci al oricărui discurs cu sens. Orice teorie ştiinţifică este numai o modalitate 'economică' de a circumscrie un număr de fapte de tipul: *dacă realizaţi cutare sau cutare acţiuni atunci veţi avea cutare sau cutare experienţe.* Susţinătorul acestei perspective nu trebuie să apere afirmaţia, greu de dovedit, că oamenii de ştiinţă sunt interesaţi numai de aplicaţii, sau numai de scopuri practice şi dezinteresaţi de cunoaştere de dragul cunoaşterii. Fenomenalistul nu trebuie să nege că vrem să cunoaştem găurile negre, că vrem să ştim dacă a existat Big Bang-ul sau care este originea lui homo sapiens etc. Vrem să ştim toate aceste lucruri nu numai pentru că ne-ar ajuta să construim maşini. Ci cunoaşterea lor, deşi este nevoie de o analiză filosofică rafinată pentru a demonstra, înseamnă cunoaşterea unui mare număr de fapte de forma *dacă realizaţi cutare sau cutare acţiuni atunci veţi avea cutare sau cutare experienţe.* Indiferent de modul în care raţiunea este interesată de ele, *faptele sunt, în ultimă analiză, instrumentale.* În acelaşi timp, nu se întrevede nici o posibilitate ca afirmaţiile despre bine să conducă la vreo predicţie de tipul: *dacă realizaţi cutare sau cutare acţiuni atunci veţi avea cutare sau cutare experienţe.* Judecăţile despre bine sau rău nu au semnificaţie cognitivă; în cuvintele empirismului logic din secolul al XX-lea, asemenea judecăţi sunt pur 'emotive'. Fenomenalismul suferă sub două aspecte. Mai întâi, afirmaţia că judecăţile despre obiectele materiale sunt traductibile prin judecăţi despre senzaţii efective sau posibile, este falsă. Investigaţia logică atentă a acestei afirmaţii, începând cu Cercul de la Viena în deceniul patru al secolului XX, i-a convins chiar pe fenomenalişti că afirmaţia

nu este întemeiată. Teoriile ştiinţifice ca întreg conduc la speranţa că avem unele experienţe dacă săvârşim anumite acţiuni; dar teza că judecăţile ştiinţei pot fi traduse în judecăţi despre ce experienţe am avea dacă am săvârşi o serie de acţiuni, a fost dovedită ca o formă inacceptabilă de reducţionism. În al doilea rând, senzaţiile sunt cu necesitate obiecte private; deşi suntem în măsură să stabilim dacă cineva are anumite senzaţii printr-o simplă întrebare, ajungem într-un fel de încurcătură epistemologică dacă se pune problema 'de unde ştim că persoana respectivă asociază aceeaşi senzaţie cu aceeaşi descripţie aşa cum faci dumneata?' În măsura în care conţinutul ştiinţei constă în predicţii cu privire la senzaţiile pe care orice fiinţă raţională le-ar avea în cazul în care ar săvârşi anumite acţiuni, cunoaşterea conţinutului presupune să fim în măsură să spunem, de pildă dacă ne-am întâlni cu vreun extraterestru, dacă are aceleaşi senzaţii cu noi etc. Din acest motiv, filosofi precum Rudolf Carnap şi Karl Popper au insistat că predicţiile observaţionale ale ştiinţei ar trebui formulate sub forma *dacă realizaţi cutare sau cutare acţiuni atunci vor avea loc cutare şi cutare evenimente observabile public* – unde atât acţiunile săvârşite cât şi evenimentele observabile care ar avea loc trebuie descrise ca obiecte 'publice', de pildă care pot fi măsurate, şi nu ca obiecte private, aşa cum sunt senzaţiile.

Pentru a rezuma, vechiul empirism sau fenomenalism pare să genereze un criteriu pentru semnificaţia cognitivă: o judecată are semnificaţie cognitivă dacă poate fi tradusă într-o judecată despre senzaţii. În schimb, fie noţiunea de 'traducere' rămâne vagă, fie chiar judecăţile ştiinţei nu satisfac criteriul de semnificaţie cognitivă. Necazul cu trasarea unei frontiere ferme între judecăţile factuale şi cele de valoare pe temeiul că numai primele sunt 'traductibile' în judecăţi despre senzaţii este că pretinsa

traductibilitate a primei clase de judecăţi nu a fost şi de fapt nu poate fi demonstrată. Reducţionismul empirist trasează o frontieră fermă între factual şi valoric, dar cu preţul modificării a ceea ce este factual.

Scopul nostru iniţial, totuşi, era să ajungem la un răspuns pentru întrebarea 'de ce este bine să fii raţional'? Primul răspuns pe care l-am examinat, şi pe care l-am respins ca prea îngust, era că raţionalitatea ne permite să descoperim legăturile între scopuri şi mijloace. A fost invocat fenomenalismul pentru că, dacă ar fi adevărat, atunci dispare conflictul aparent dintre a fi interesat într-o teorie ştiinţifică numai de dragul consecinţelor instrumentale şi a fi interesat într-o teorie pentru a învăţa ceva despre procesele naturale. Conflictul dintre interesul instrumental şi cel pur teoretic ar putea fi, într-un sens, încheiat. Desigur, s-ar menţine alte diferenţe între aceste interese; dar chiar teoreticienii puri ar fi interesaţi în fapte care, în ultimă analiză logică, sunt de natură instrumentală. *Orice* cunoaştere care şi-ar merita numele s-ar dovedi cunoaştere despre raportul dintre mijloace şi scopuri. Atunci când suntem interesaţi de conexiunea dintre mijloace şi scopuri, datorită faptului că dorim să ştim cum să ajungem la un scop, am numi acest interes 'practic', iar atunci când dorim să ştim raportul între mijloace şi scopuri numai din curiozitate, avem de-a face cu un interes teoretic. Această tentativă de a reduce toate judecăţile ştiinţei la judecăţi de forma *dacă realizaţi acţiunea A atunci veţi obţine rezultatul B* eşuează, aşa cum am arătat.

Obiecţia lui Carnap şi Popper că judecăţile de observaţie ale ştiinţei sunt exprimabile în limbajul obiectelor fizice şi nu în cel al senzaţiilor, este de bună seamă corectă dacă o considerăm ca generalizare a practicii ştiinţifice. Pusă în relaţie cu absolutul epistemologic, obiecţia devine decisivă

cel puţin pentru că, dacă nu permitem nici unei judecăţi de observaţie să vorbească despre senzaţii, atunci introspecţia este exclusă dintre metodele ştiinţei. Deşi mulţi psihologi ar fi de acord cu aceasta, în fapt unii psihologi, fără a ţine seama de dogmele filosofice şi psihologice, au săvârşit experimente care presupun cel puţin în parte raportarea introspecţională. De fapt, Carnap nu era atât de dogmatic în respingere precum unii psihologi behaviorişti; el ar fi permis raportarea la senzaţii sub condiţia să fie construită nu ca raport de observaţie ci prin intermediul unor date comportamentale, unele fiind implicate chiar în emiterea raportului verbal. Dar ce înseamnă a analiza acceptarea unui raport introspectiv ca o 'inferenţă din comportamentul verbal' nu este deloc clar. Un raport care nu este despre senzaţii ci despre obiecte fizice, cum ar fi 'în faţa mea se află o masă', este acceptat numai când avem o teorie conform căreia persoana care face un asemenea raport este în situaţia de a observa faptul respectiv. În acest sens, ca parte a cerinţelor de coerenţă pentru imaginea noastră asupra lumii, ar trebui ca observaţiile să fie, la rândul lor, explicabile teoretic; dacă cineva afirmă că într-un anume loc se află o masă prin *clarviziune* nu acceptăm un asemenea 'raport observaţional', deoarece nu este coerent cu corpul teoretic acceptat. În acest sens *orice* raport observaţional are o componentă care ar putea fi descrisă ca 'inferenţială'. Pe de altă parte, când, să spunem, un doctor acceptă raportul unui pacient că simte dureri, este greu să ştim pe baza cărei 'teorii ştiinţifice' doctorul inferează că pacientul simte durerea pe care o raportează; dacă presupunerea generală că oamenii sunt în măsură să spună dacă au o durere este considerată teorie, atunci presupunerea generală că oamenii sunt în măsură să spună dacă există o masă în faţa lor contează ca teorie, deoarece este dificil să vedem vreo diferenţă metodologică între acceptarea spuselor

cuiva cu privire la o masă şi acceptarea spuselor cuiva cu privire la durere.

Carnap şi Popper ar răspunde că diferenţa metodologică este că prima judecată este public verificabilă, în vreme ce a doua nu este; dar ei exagerează măsura în care rapoartele observaţionale cu privire la obiectele fizice sunt întotdeauna verificabile public. Multe dintre asemenea rapoarte sunt făcute cu ajutorul unor instrumente care necesită o pregătire prealabilă. (Este ştiut faptul că a învăţa să vezi printr-un microscop puternic cere o pregătire de specialitate şi o îndemânare care nu este la îndemâna oricui.) A accepta că verificabilitatea publică este o parte a raţionalităţii reprezintă o dogmă epistemologică. A face din aceasta o parte a definiţiei raţionalităţii este foarte convenabil; nu mai este nevoie să se aducă argumente în această privinţă. Poate că argumentul este, la urma urmei, că orice nu este public verificabil poate fi obiect de disensiune care, dacă nu poate fi tranşată, nu se poate afirma ce este corect şi ce este eronat. Dar astfel s-ar adopta teza majorităţii, adică ideea că verificabilitatea raţională înseamnă verificabil cu satisfacerea unei majorităţi covârşitoare.

Faptul că empirismul logic era o expresie sofisticată a tendinţelor culturale instrumentaliste şi a tezei majorităţii devine evident, cred, în ultima perioadă a mişcării. Deşi empiriştii logici au abandonat fenomenalismul încă din 1936, în următorii 20 de ani, când mişcarea începuse să slăbească, filosofii empirişti logici ai ştiinţei vorbeau despre 'scopul ştiinţei' şi tindeau să-l identifice cu *predicţia* (cu unele adausuri şi calificative pe care le voi discuta imediat). Teza că scopul ştiinţei este predicţia a reperzentat ideea fundamentală a pozitivismului odată cu scrierile lui Auguste Comte. Cum am văzut, aici sunt prezente cel

puţin un tip de raţiuni filosofice cât timp fenomenalismul era la modă; deoarece atunci se putea argumenta că toate judecăţile cu semnificaţie cognitivă sunt predicţii deghizate. Reapariţia acestei doctrine după dispariţia fenomenalismului este ca supa reîncălzită. A spune că scopul ştiinţei îl reprezintă succesul predicţiilor (sau succesul predicţiilor plus ceva numit 'simplitate') pare foarte aproape de a spune că ştiinţa este condusă numai de obiective practice; dar nici un filosof nu ar dori să susţină o asemenea consecinţă. Într-adevăr, filosofii care apără o poziţie pur instrumentalistă privind ştiinţa nu o fac pentru că ar fi ei înşişi fanatici ai practicii, sau oameni mărginiţi care nu pot aprecia frumuseţea ştiinţei abstracte de dragul ei, ci mai degrabă pentru că, prin identificarea 'semnificaţiei cognitive' cu valoarea pentru predicţii, ar putea să evite orice formă de obscurantism şi metafizică. 'Metafizica' era pentru aceşti filosofi un alt nume pentru diferite tipuri de speculaţie transcendentală pe care ei voiau să o evite.

Sugerez că apariţia unei tendinţe filosofice care era hipnotizată de succesul ştiinţei în aşa fel încât nu putea concepe posibilitatea cunoaşterii şi raţiunii în afară de ceea ce le plăcea să numească ştiinţific, este ceva de aşteptat, dat fiind enormul prestigiu al ştiinţei în cultură şi declinul prestigiului eticii absolute, al metafizicii transcendentale şi al religiei. Sugerez, de asemenea că înaltul prestigiu al ştiinţei în cultură este dat îndeosebi de succesul instrumental al ştiinţei, împreună cu faptul că ştiinţa pare liberă de dezbateri interminabile şi irezolvabile, aşa cum sunt religia, etica şi metafizica.

Dacă, totuşi, filosofii profesionişti care, cum am văzut, raţionalizează tendinţa instrumentalistă în cultură, nu sunt ei înşişi înclinaţi spre vulgarizare sau persoane pur practice, nu este surprinzătoare înclinaţia lor spre

deschiderea 'scopului ştiinţei' şi spre alte orizonturi decât predicţiile de succes. Astfel ajungem la celelalte scopuri puse în evidenţă de empirismul logic între anii 1940–1950: descoperirea legilor, retrodicţia etc. (care vizează evenimentele din trecut) şi descoperirea de 'explicaţii' prin care aceşti autori înţelegeau de fapt deducerea predicţiilor şi retrodicţiilor din legi.

Este ciudat ce s-a întâmplat. Pentru a face clar că ştiinţa este interesată de descoperirea legilor naturii de dragul acestora şi nu pentru predicţii prin intermediul lor, aceşti autori înlocuiesc formula 'scopul ştiinţei constă în succesul predicţiilor' cu o *listă*. Lista este de fapt deschisă-închisă: legile naturii sunt astfel circumscrise încât să includă nu numai legile naturii în sens strict, cum ar fi judecăţile care nu pot fi fizic false ci şi aşa-numitele legi ale evoluţiei, care sunt, de fapt, descrieri ale tendinţelor generale care pot, la un moment dat, să genereze viaţa inteligentă, şi chiar judecăţi despre dispoziţii ale grupurilor şi organismelor. Este, de bună seamă, corect a spune că oamenii de ştiinţă încearcă să descopere 'legile naturii' inclusiv generalizările fizice contingente care au loc pentru lungi perioade de timp şi care au o largă putere explicativă, cum ar fi cele pe care se bazează teoria evoluţiei sau cele pe care le au în vedere ştiinţele economice, cât şi descoperirea adevărurilor semnificative cu privire la dispoziţiile grupurilor şi organismelor individuale şi cercetarea organizării acestora într-o structură deductivă (şi inductivă). Dar de ce este nevoie de o listă?

Prin alcătuirea listei se urmăreşte surprinderea tuturor acelor tipuri de adevăruri pe care le au în vedere oamenii de ştiinţă, încât să conţină ştiinţa fizicii, dar să excludă aspectele care nu sunt cognitiv semnificative. Vechea investigaţie pentru 'criteriul semnificaţiei cognitive', cum ar fi 'o propoziţie are sens dacă şi numai

pe care oamenii de ştiinţă sunt interesaţi să le descopere, cu siguranţă nu sunt prevăzute aici judecăţile istoricului, ale psihiatrului şi sociologului; pe de altă parte, dacă termenii listei sunt construiţi astfel încât să includă fără nici o dificultate judecăţile făcute de istoric (şi judecăţile descriptive în limbajul psihologiei), atunci lista devine fără valoare. În orice caz, în absenţa oricărei explicaţii epistemologice asupra faptului de ce judecăţi *de aceste feluri* şi numai acestea ar trebui să fie capabile de verificare raţională, o asemenea listă nu ar fi decât o simplă ipoteză asupra limitelor cercetării raţionale. Iar o simplă ipoteză, în formă de listă sau într-o altă formă, nu ar avea forţa unor 'criterii pentru semnificaţia cognitivă' pe care empirismul logic o dorea.

Fetişismul 'metodei'

Dacă răspunsul la întrebarea 'de ce este bine să fi raţional?' nu poate fi pur şi simplu că raţionalitatea permite atingerea unor scopuri practice sau că raţionalitatea permite descoperirea legăturilor dintre mijloace şi scopuri, am putea lua în considerare o altă variantă de răspuns, care a avut o susţinere considerabilă la diferite momente. Mulţi filosofi au crezut că ştiinţa urmează o *metodă* specifică; dacă există o metodă cu proprietatea că folosind-o, ajungem la adevăr şi dacă adevărul nu poate fi descoperit prin nici o altă metodă şi dacă explicaţia succesului extraordinar al ştiinţei şi persistenţa controverselor în alte domenii diferite de ştiinţă este că ştiinţa şi numai ea foloseşte sistematic această *metodă*, atunci poate că raţionalitatea, în măsura în care există, ar trebui să fie identificată cu posedarea şi folosirea metodei. Răspunsul la întrebarea 'de ce este bine

să fii raţional?' ar fi: dacă eşti raţional ajungi la adevăr (oricare ar fi adevărurile care ne interesează) în vreme ce dacă nu eşti raţional nu ai şanse reale să descoperi adevărul. Precum perspectiva instrumentalistă, acest punct de vedere a cunoscut, în decursul istoriei filosofiei, momente de creştere, stagnare şi declin. De la publicarea *Logicii* lui Mill, în 1840, până la *Fundamentele logice ale probabilităţii* a lui Carnap, filosofi influenţi ai ştiinţei au continuat să creadă că ştiinţele empirice utilizează metoda formală ('*logica inductivă*') şi că s-ar putea ajunge la explicitarea acestei metode, la o formalizare a logicii inductive comparabilă cu formalizarea logicii deductive care a fost reuşită în urma operei lui Frege. Dacă s-ar fi descoperit o asemenea metodă, atunci chiar dacă nu este o *dovadă* prin ea însăşi că s-ar suprapune raţionalităţii, încă esenţa demonstraţiei ar însemna foarte mult împotriva celor care afirmă că există adevăruri care pot fi justificate şi arătate ca fiind raţional acceptabile prin alte mijloace.

Conform celei mai influente şcoli, aşa-numita şcoală 'bayesiană', caracterul general al metodei inductive pe care filosofii doresc să o formalizeze este următorul: presupunem că limbajul ştiinţei a fost formalizat şi că oamenii de ştiinţă au la dispoziţie un anumit număr de observaţii exprimabile prin 'propoziţii de observaţie' în acest limbaj formalizat. Presupunem, de asemenea, că sunt luate în considerare ipoteze diferite care sunt exprimate prin formule ale acestui limbaj. Problema logicii inductive este de a defini o 'funcţie de confirmare', care este o funcţie de probabilitate care va determina o probabilitate matematică relativă la evidenţa observaţională pentru fiecare dintre ipoteze sau, în altă terminologie, 'gradul de confirmare' pe care evidenţa o are pentru fiecare dintre ipotezele rivale. De regulă, se presupune că se cunoaşte

probabilitatea evidenţei dacă o ipoteză ar fi adevărată; aceasta se numeşte 'probabilitatea directă', respectiv probabilitatea evidenţei *dându-se* ipoteza. Se doreşte să se calculeze aşa-numita 'probabilitate inversă', respectiv, probabilitatea ipotezei în raport cu evidenţa dată. Teorema lui Bayes dă posibilitatea acestui calcul ca o funcţie simplă de probabilităţile directe şi de alte probabilităţi, numite 'probabilităţi anticipate' pentru ipotezele alternative, care sunt probabilităţi sau 'grade subiective de încredere' presupuse de oamenii de ştiinţă anterior raportării la evidenţa observaţională.

Probabilităţile directe sunt uşor de calculat în situaţii precum: (a) ipoteza implică evidenţa (în acest caz 'probabilitatea directă' este *unu*); (b) ipoteza este statistică sau stochastică având conţinutul că evidenţa ar trebui să apară cu o anumită probabilitate *r*. Teorema lui Bayes se confruntă cu o dificultate atât de serioasă încât numeroşi filosofi şi statisticieni nu sunt de acord privind importanţa şi utilitatea teoremei în ce priveşte teoria confirmării: pentru măsurarea probabilităţii directe este nevoie de 'gradele subiective de încredere', în terminologia lui De Finetti şi Savage.

Să ne mulţumim pentru moment cu presupunerea că există 'probabilităţi directe', care pot fi calculate. Într-un asemenea caz, metoda este una pur formală; am putea programa o maşină să calculeze gradele în care ipotezele sunt susţinute de evidenţă. Numai că 'intrările' ar include nu numai probabilităţile directe ci şi probabilităţile anticipate în contextul dat. Dacă le considerăm pe acestea ca fiind convingerile anterioare ale oamenilor de ştiinţă despre lume, după cum sugerează termenul 'funcţie subiectivă de probabilitate', atunci una dintre intrări este chiar o mulţime de convingeri (sau grade de convingeri) despre

lume. Acesta este modul în care mulţi filosofi ai ştiinţei de astăzi înţeleg problema; din ce în ce mai mult se ajunge să se creadă că nu este posibil să se traseze o frontieră netă între *conţinutul* şi *metodele* ştiinţei; că metoda ştiinţei, de fapt, se modifică în timp după conţinutul ştiinţei. Teorema lui Bayes, dacă într-adevăr surprinde logica teoriei confirmării, generează o modalitate prin care dependenţa metodei de conţinutul ştiinţei poate fi formalizată prin intermediul funcţiei probabilităţii directe.

Pentru a pune problema într-o formă mai abstractă, am putea spune că fetişismul metodei presupune că raţionalitatea este *indivizibilă*. Dar teorema lui Bayes arată că lucrurile nu stau aşa; ci putem separa raţionalitatea, chiar în domeniul ştiinţei, îndeosebi când avem de-a face cu teorii pentru care probabilităţile directe pot fi analizate în componente: o componentă *formală*, care poate fi schematizată matematic şi programată pe un computer şi o componentă *neformală* care nu poate fi formalizată şi care depinde de schimbările convingerilor oamenilor de ştiinţă. Acum, ar fi bine dacă partea formală a raţionalităţii ar fi suficientă pentru a garanta rezultatele aşteptate. Dacă oamenii de ştiinţă efectuează observaţiile cu multă grijă, au la dispoziţie suficiente observaţii şi calculează gradele de întemeiere conform teoremei lui Bayes atunci, probabil că ar cădea de acord, chiar dacă nu au fost la început, că neînţelegerile provin de la diferenţa dintre convingerile lor subiective. Dar sunt două aspecte care nu se potrivesc cu acest scenariu.

Mai întâi, chiar dacă am putea arăta că pe termen lung 'funcţia probabilităţii directe' se anulează, sau că oamenii de ştiinţă cu convingeri iniţiale diferite ajung la acord cu cât utilizează mai mult teorema lui Bayes, ar fi încă nevoie ca această convergenţă să se producă suficient

de repede. Dacă oamenii de ştiinţă cu diferite opinii iniţiale nu ajung la acord până la momentul la care fenomenul anticipat ar trebui să aibă loc sau timp de milioane de ani, atunci, pe durată scurtă, faptul că este matematic sigur că vor ajunge la consens nu mai este util; problema cu întemeierea într-un timp lung este că acest timp ar putea fi prea lung. În faimoasele cuvinte ale lui John Maynard Keynes 'pe termen lung toţi vom fi morţi'. Al doilea aspect este că, de fapt, diferenţele privind convingerile anterioare pot genera diferenţe violente privind gradul de întemeiere a ipotezelor, care pot fi considerate chiar iraţionale.

Pentru a formula altfel, un om de ştiinţă va presupune numai grade de întemeiere a ipotezelor care i se par 'rezonabile' dacă porneşte de la funcţii de probabilitate anticipată rezonabile. Dacă cineva se supune numai părţii formale a descrierii raţionalităţii, dacă rămâne consistent şi calculează gradele de întemeiere conform cu teorema lui Bayes, dar funcţia probabilităţii anticipate este extrem de 'nerezonabilă', atunci judecăţile sale despre măsura în care sunt întemeiate diferite ipoteze de către evidenţă vor fi 'iraţionale'. Raţionalitatea formală, angajamentul faţă de partea formală a metodei ştiinţifice, nu garantează atingerea raţionalităţii reale.

Acest adevăr apare mai degrabă şocant. Arthur Burks a arătat că există chiar 'funcţii de probabilităţi anticipate contrainductive'. Adică, există un tip logic posibil de asemenea funcţii astfel încât, dacă un om de ştiinţă le acceptă, atunci cu cât există mai multă evidenţă pentru o ipoteză (folosim termenul de 'mai multă evidenţă' pe baza judecăţilor inductive normale), omul de ştiinţă va atribui o probabilitate tot mai mică ipotezei pe termen foarte lung.

O cale de a depăşi dificultăţile semnalate ar putea fi îmbogăţirea componentei formale a metodei ştiinţifice prin admiterea unei mulţimi de reguli pentru a determina care probabilităţi anticipate sunt rezonabile (de aici încolo, eu mă voi referi la funcţia de probabilităţi anticipate prin 'anticipate') şi care nu sunt rezonabile. Dar nu pare să existe vreo motivaţie potrivită pentru convingerea că ar exista o asemenea mulţime de reguli prin ce s-ar putea deosebi între anticipate rezonabile şi nerezonabile, care ar fi mai simple decât descrierea completă a psihologiei şi idealului raţional al umanităţii în întregul ei. Speranţa într-o metodă formală, posibil de izolat pornind de la judecăţile actuale acceptate de oameni cu privire la conţinutul ştiinţei (adică despre natura lumii) şi la valorile umane, pare a fi imposibil de împlinit. Chiar dacă lărgim noţiunea, aşa încât formalizarea psihologiei şi a idealului raţional uman să conteze ca metodă, nu există nici un temei să se creadă că o astfel de 'metodă' ar fi independentă de judecăţile privind estetica, etica sau altfel de judecăţi. Motivaţia de ansamblu pentru a crede că metoda ştiinţifică nu s-ar aplica la, sau nu ar presupune convingeri de natură etică, estetică etc., ar fi de fapt, convingerea că metoda ştiinţifică este o metodă *formală*.

Discuţia de aici a depins de supoziţia asupra corectitudinii unei abordări particulare a formalizării metodei ştiinţifice, aşa-numita abordare bayesiană. Dar aceleaşi probleme apar în oricare alte abordări pe care le-am lua în considerare. Chiar dacă cineva ar încerca să izoleze o mică parte a metodei inductive care nu ar fi atât de puternică precum confirmarea teoriilor şi care s-ar afla mai mult pe linia deschisă de Bacon în înţelegerea inducţiei, adică chiar dacă se încearcă izolarea unei metode a confirmării adevărului unor generalizări simple prin examinarea a suficient de multe instanţe şi 'proiecţii', s-ar pune probleme

similare. Nelson Goodman[5] a arătat că nici o regulă *pur* formală pentru proiecţia inductivă nu poate fi liberă de inconsistenţă. Înaintea unei reguli formale ar fi de ajutor separarea predicatelor considerate 'proiectibile' ale limbajului de cele considerate ca 'neproiectibile'. Faptul că şi cea mai elementară inducţie presupune o partiţie (respectiv o diviziune a vocabularului în părţi proiectibile şi neproiectibile) care este neformală, din nou sprijină concluzia discuţiei despre teorema lui Bayes, că nu se poate trasa o delimitare fermă între convingerile actuale ale oamenilor de ştiinţă şi metoda ştiinţifică. Goodman arată tocmai acest lucru prin inventarea predicatului 'verbastru' care se aplică lucrurilor numai în cazul în care sunt *observate înainte de anul 2 000 şi sunt verzi şi dacă nu sunt observate înainte de anul 2 000 şi sunt albastre.* Înainte de anul 2 000, orice este examinat şi este văzut ca fiind verde este verbastru. Orice regulă formală de proiecţie potrivit căreia, după ce am examinat un anumit număr de lucruri, să spunem smaralde, cu proprietatea *P* putem să inferăm că 'toate smaraldele au *P*' ne-ar permite să facem inferenţa contradictorie că 'toate smaraldele sunt verzi' şi 'toate smaraldele sunt verbastre'. Pe de altă parte, Goodman arată în mod convingător că toate încercările de a determina printr-o regulă asemenea predicate bizare cum este 'verbastru' pe un temei pur formal nu pot fi duse la bun sfârşit[6].

---

[5] Vezi lucrarea sa *Fact, Fiction and Forecast*, Hackett (1977), publicată prima dată în 1954.

[6] Soluţia lui Goodman este de a considera forma *plus* istoria proiecţiei predicatelor prezente în inferenţă (împreună cu unele aspecte corelate). Potrivit propunerii lui Goodman ar urma că o cultură care are *întotdeauna* proiectate asemenea predicate ciudate precum celebrul predicat 'verbastru' ar fi perfect justificată să procedeze astfel –

Există o legătură strânsă între dificultatea pe care o sesizează Goodman în cazul inducţiei baconiene şi cele relatate despre teorema lui Bayes. Să presupunem că un om de ştiinţă are de ales între două ipoteze (cândva înainte de anul 2 000): 'toate smaraldele sunt verzi' şi 'toate smaraldele sunt verbastre'. Să presupunem că evidenţa relevantă este că au fost examinate numeroase smaralde şi au fost găsite verzi (astfel încât ele sunt şi verbastre). Dacă omul de ştiinţă calculează gradul de întemeiere pentru cele două ipoteze utilizând metoda lui Bayes, rezultatul la care ajunge depinde de convingerile anticipate. Dacă convingerile anterioare ale cuiva conduc spre o probabilitate mai mare pentru 'toate smaraldele sunt verzi' decât pentru 'toate smaraldele sunt verbastre', atunci se întăreşte proiecţia pentru 'verde' în detrimentul celei pentru 'verbastru'. Din punct de vedere bayesian necesitatea unei decizii privind care ipoteze sunt proiectibile înainte de a face o inducţie este doar un caz special al necesităţii de a estima convingerile anterioare.

Karl Popper a sugerat că ar trebui acceptată *cea mai falsificabilă* dintre ipotezele alternative; dar prin măsurarea pur formală a falsificabilităţii se ajunge la rezultate diferite în funcţie de predicatele limbii alese ca primitive. Dacă cineva crede despre omul de ştiinţă, aidoma lui Popper, că încearcă să determine *cea mai falsificabilă dintre ipotezele asupra cărora încă nu s-a decis* sau crede într-un mod mai convenţional că încearcă să calculeze *gradele de întemeiere ale ipotezelor*, este încă necesară nevoia de elemente informale, corespunzător deciziei goodmaniene că anumite predicate sunt proiectibile, iar altele nu, sau corespunzător acceptării probabilităţilor anticipate bayesiene.

---

inferenţele lor ar fi inductiv valide! În timp ce sunt de acord cu Goodman că a ţine seama de practica anterioară este un principiu important al ştiinţei, versiunea lui Goodman a acestui principiu este prea simplă şi relativă.

La acest moment cititorul poate fi mirat, dacă nu există o metodă ştiinţifică sau dacă metoda, în măsura în care poate fi formalizată, depinde de elemente care nu pot fi formalizate, atunci cum se explică succesul ştiinţei? Nu se poate nega că ştiinţa a cunoscut un succes deosebit. Credem că motivul acestui succes trebuie să aibă de-a face cu diferenţa între modul în care oamenii de ştiinţă procedează pentru a dobândi cunoaştere şi modul în care se ajungea la cunoaştere în perioada preştiinţifică. Oare acesta din urmă este eronat în totalitate? Răspunsul este nu. Alternativele între care avem de ales nu sunt că ştiinţa are succes pentru că urmează un fel de algoritm riguros, pe de o parte şi că ştiinţa are succes din pură întâmplare. Începând cu secolul al XV-lea şi atingând un fel de maxim în secolul al XVII-lea, oamenii de ştiinţă şi filosofii au început să formuleze o nouă mulţime de maxime metodologice care nu sunt atât de riguroase ca şi regulile formale; ele cer o raţionalitate informală, cum ar fi inteligenţa şi simţul comun; dar neîndoielnic sunt adaptate cercetării ştiinţifice. Pe scurt, există o metodă ştiinţifică; dar aceasta presupune noţiuni anterioare de raţionalitate[7]. Nu este o metodă *nouă* care poate servi ca definiţie a raţionalităţii.

Unul dintre cei mai însemnaţi metodologi din secolul al XVII-lea a fost fizicianul Boyle. Anterior secolului al XVII-lea, fizicienii nu distingeau prea bine între a efectua un experiment şi a descrie experimente mentale care ar confirma teoriile în care ei credeau deja, mai mult sau mai puţin întemeiat. Mai mult, fizicienii nu simţeau nevoia de a face publică descrierea experimentelor care dădeau greş. Pe scurt,

[7] Mill însuşi acceptă aceasta (chiar dacă indirect) când scrie că nu ne putem aştepta ca metoda inductivă să funcţioneze 'împreună cu prostia universală' (*Utilitarism*, Cap. 2).

experimentele erau înţelese mai degrabă ca *ilustraţii* pentru doctrinele în care credeau prin întemeiere deductivă sau a priori; nu ca evidenţe împotriva teoriilor. Boyle a scris manuale privind procedeele experimentale, a arătat necesitatea unei distincţii clare între experimentele mentale şi cele reale, ca şi necesitatea de a da o descriere completă pentru toate experimentele efectuate, inclusiv a celor care au eşuat. Boyle era un discipol al filosofului Francis Bacon şi era, fără îndoială, sub influenţa regulilor inductive ale lui Bacon; de fapt, se poate ca instrucţiunile specifice date de Boyle să fi fost mai importante sau la fel de importante ca şi procedeele inductive mult mai schematice şi abstracte ale lui Bacon pentru cursul pe care l-a luat cercetarea fizică.

Cotitura de la teoriile *a priori*, la încercarea de a testa teorii prin derivarea unor concluzii efectuând experimente, a însemnat o importantă schimbare metodologică. După cum am văzut, nu putem totuşi identifica *raţionalitatea* cu *încrederea într-o teorie numai pentru că aceasta este sprijinită de experimente efectuate migălos*, pentru că, chiar în cazul ştiinţei, nu este posibil întotdeauna să se realizeze experimente controlate. Uneori, trebuie să avem încredere în observaţii pasive mai degrabă decât într-o intervenţie activă presupusă de termenul 'experiment'. Pe lângă asta, aşa cum s-a văzut, chiar atunci când experimentele sunt realizate cu multă grijă pentru a selecta între mai multe teorii concurente, estimarea gradului în care rezultatele experimentale susţin aceste teorii alternative este încă grevată de aspecte neformale.

Împotriva afirmaţiilor de mai sus, Karl Popper a argumentat de mai multe ori că există o metodă ştiinţifică distinctă, că poate fi afirmată şi ar trebui să avem încredere numai în ea pentru a descoperi natura acestei lumi.

Popper admite, cu toate acestea, că există noţiuni de raţionalitate care sunt mai largi decât raţionalitatea ştiinţifică şi care se aplică atunci când este vorba de decizii etice.

Popper, în cartea *Logica cercetării*, precum şi în alte publicaţii, argumentează că metoda conduce la impunerea teoriilor 'puternic falsificabile'; a teoriilor care permit predicţii riscante. În acest caz, teoriile sunt supuse unor teste în faţa cărora supravieţuieşte numai una. După aceea, acceptăm ipoteza care rămâne şi aceasta funcţionează o vreme până ce întreaga procedură se repetă. Deoarece eliminarea teoriilor cu excepţia uneia este realizată prin mijloace deductive – o teorie este eliminată când implică o predicţie care este dovedită ca falsă – nu este cerută utilizarea teoremei lui Bayes şi nici estimarea gradelor de confirmare.

O problemă care afectează strategia lui Popper este imposibilitatea de a testa toate teoriile puternic falsificabile. De exemplu, teoria că, dacă ţinem un sac de făină pe cap şi lovim masa de 99 de ori va apare un demon, este puternic falsificabilă, dar cu siguranţă nu o pot testa. Chiar dacă cineva ar dori, într-o viaţă de om, ba chiar în decursul existenţei speciei umane tot nu s-ar putea testa toate teoriile rivale. Din motive lesne de înţeles, atunci, este nevoie să alegem, pe temeiuri metodologice, un număr mic de teorii pe care să le testăm efectiv; asta înseamnă că şi în strategia popperiană apare ceva analog selecţiei anterioare. După cum am remarcat, calculul popperian al gradelor de falsificabilitate depinde de predicatele pe care le considerăm primitive, aşa încât noţiunea de falsificabilitate necesită o decizie anterioară de tipul celei puse în evidenţă de Goodman cu privire la predicatele proiectibile. Să renunţăm la aceste aspecte tehnice care nu sunt de mare interes în ce priveşte discuţia noastră. Chiar dacă metoda lui Popper este incompletă şi trebuie suplimentată cu o metodă mai intuitivă pe care nu suntem în măsură să o formalizăm deocamdată, oare nu descrie o

condiţie necesară, chiar dacă nu suficientă, pentru raţionalitatea ştiinţifică? Oare supravieţuirea în faţa testelor popperiene nu ar fi o condiţie necesară pentru acceptarea unei teorii ştiinţifice? Testele lui Popper pot presupune o selecţie anterioară a teoriilor care trebuie testate, selecţie care este neformală şi pentru care nu avem un algoritm; calculul privind teoriile cele mai falsificabile presupune decizii neformale pentru care nu avem algoritm; dar se poate totuşi insista că nici o teorie să nu fie acceptată decât în condiţiile în care o mulţime de teorii a fost mai întâi selectată ca *intuitiv* 'puternic falsificabilă', iar dintre ele toate, cu excepţia celei pe care o acceptăm, au fost respinse de experimente efectuate cu grijă. Pe scurt, de ce să nu dăm unui om de ştiinţă următorul sfat: procedează aşa cum spune Popper şi, acolo unde metodele lui Popper nu pot fi formalizate, ai încredere în propria intuiţie pentru modul în care ar trebui interpretate maximele lui Popper? S-ar putea ca metoda popperiană, parţial vagă şi neformală, să nu se suprapună nici peste noţiunea de raţionalitate *ştiinţifică* nici peste cea de *raţionalitate cognitivă*, adică s-ar putea ca judecăţile să nu fie asertabile în mod justificat, sau să nu fie acceptabile raţional, dacă şi numai dacă sunt implicate de o teorie care poate fi acceptată prin teste Popper? Răspunsul este că o asemenea concepţie despre raţionalitate este prea îngustă chiar şi pentru ştiinţă, deoarece respinge una dintre cele mai de succes şi admirate teorii ştiinţifice, şi anume teoria evoluţiei prin selecţie naturală a lui Darwin. Aceasta este o consecinţă pe care Popper însuşi ar fi dispus să o accepte în cele din urmă, dar, cu siguranţă, comunitatea ştiinţifică o respinge. Teoria selecţiei naturale nu este înalt falsificabilă; nu implică predicţii definite astfel încât, dacă ele eşuează, teoria să fie respinsă. Acceptăm teoria selecţiei naturale nu pentru că a supravieţuit testelor popperiene, ci pentru că produce *explicaţii plauzibile* pentru un număr foarte mare de date,

pentru că a fost utilă în dezvoltarea altor teorii şi pentru că este legată de succesele din genetică, biologia moleculară etc. şi deoarece teoriile alternative fie au fost falsificate, fie par neplauzibile în lumina fundalului cunoaşterii. Pe scurt, acceptăm teoria evoluţiei prin selecţie naturală a lui Darwin potrivit mecanismului pe care Peirce îl numea 'abducţie' sau cum se numeşte mai recent 'inferenţa la cea mai bună explicaţie'. Acesta este chiar tipul de inferenţă pe care Popper intenţiona să-l elimine din ştiinţă; dar oamenii de ştiinţă nu s-au lăsat convinşi că ar trebui să renunţe la teoriile care nu sunt puternic falsificabile dacă aceste teorii oferă explicaţii pentru o mare varietate de date şi dacă produc explicaţii plauzibile în domeniul lor. Într-adevăr, aşa cum am subliniat într-o altă lucrare, Popper exagerează măsura în care chiar teoriile fizicii clasice sunt puternic falsificabile[8].

Să slăbim descrierea metodei ştiinţifice până la nivelul 'inferenţei la cea mai bună explicaţie' ca formă legitimă pentru inferenţa ştiinţifică, chiar dacă 'cea mai bună explicaţie' are în vedere cazuri care nu sunt puternic falsificabile în sensul lui Popper. Metoda ştiinţifică devine astfel periculos de vagă[9]; dar era de aşteptat să se întâmple aşa dacă ţinem seama de rezultatele logicii inductive prezentate anterior. Oare 'metoda ştiinţifică', descrisă atât de vag, este acum exhaustivă? Oare o asemenea descriere vagă să fie suficientă odată ce nici o judecată de valoare nu poate fi verificată sau confirmată printr-o asemenea metodă? Dacă metoda ştiinţifică se reduce la 'faceţi experimente şi

[8] Vezi 'The Corroboration of Theories', în volumul meu *Mathematics, Matter and Method*.

[9] Ca alternativă am putea restrânge termenul 'metodă ştiinţifică' pentru a se referi la aplicaţiile intenţionate ale procedurii experimentale maximale, aşa cum recomand în *Meaning and the Social Sciences*, şi să încetăm în tentativa de a-l face atât de elastic încât să cuprindă întreaga 'cunoaştere'.

observaţii cât de atent puteţi, iar pe baza acestora faceţi inferenţe către cea mai bună explicaţie şi eliminaţi teoriile care pot fi falsificate prin experimente cruciale' atunci este imposibil de stabilit ce *nu poate fi* verificat printr-o metodă atât de vag descrisă. Să presupunem, de pildă, că vrem să verificăm judecata 'John este un om rău'. S-ar putea argumenta astfel: 'Au fost observate următoarele fapte: John nu este politicos, John este foarte egoist şi John este o persoană foarte crudă. Cineva care nu este politicos, care este egoist şi crud este *prima facie* o persoană rea; prin urmare, John este un om rău'. Există două aspecte cu privire la care un susţinător al punctului de vedere că 'judecăţile de valoare' nu pot fi 'verificate ştiinţific' ar putea obiecta la acest argument. El ar putea obiecta faţă de ultimul pas, respectiv pasul de la *John nu este politicos, este egoist şi crud, la John este un om rău*. Să admitem că acesta este un pas *conceptual* care afirmă că există o legătură conceptuală între a fi crud, a nu fi politicos şi a fi egoist, şi a fi rău din punct de vedere moral[10]. De bună seamă că, dacă nu există nici o legătură conceptuală între predicatele morale, atunci pasul este nevalid; dar de ce să nu credem că există legături conceptuale între predicate morale? S-ar putea argumenta că folosirea paşilor descrişi

[10] Faptul că un adevăr sau o inferenţă este de tipul pe care îl numesc 'conceptual' nu înseamnă că este *lingvistic* pur (adevărat în virtutea convenţiilor lingvistice *arbitrare*). Filosofi de tendinţe diferite au observat cum conceptele, faptele şi observaţiile sunt interdependente. După cum am remarcat în Capitolul 6, conceptele sunt modelate prin observaţie sau intuiţie şi la rândul lor, modelează observaţia sau intuiţia. Ţinând seama de acestea, inferenţa într-un text implicând 'binele' este analogă cu următoarea inferenţă care apelează la 'conştiinţă'. 'John vorbeşte inteligent, acţionează potrivit şi este răspunzător de ce se petrece; deci John are conştiinţă'. Legătura conceptuală este că 'a vorbi *inteligent*', 'a acţiona *potrivit*', '*a răspunde* pentru ce se petrece' sunt *prima facie* motive pentru a atribui conştiinţă în acelaşi fel în care bădărănia, egoismul şi cruzimea sunt *prima facie* motive pentru a atribui răutatea morală.

drept 'conceptuali' într-un argument este neştiinţifică; dar nu se poate respinge că există asemenea inferenţe în orice ştiinţă. De exemplu, când inferez de la descrierea sistemului solar a lui Newton la judecata 'atracţia gravitaţională a Lunii este cauza mareelor', utilizez cunoaşterea neformală că există o legătură conceptuală între judecăţile despre forţe şi judecăţile de forma '*A* cauzează *B*'. Cuvântul 'cauză' nu apare în descrierea lui Newton a sistemului solar şi a mareelor; dar ştiu că forţa gravitaţională pe care *A* o exercită asupra lui *B* poate fi descrisă drept cauzată de (masa lui) *A* în virtutea interpretării teoriei lui Newton.

De bună seamă, dacă admitem că metoda ştiinţifică se reduce la 'inferenţa la cea mai bună explicaţie' din 'judecăţi de observaţie' care sunt *într-un limbaj neutru valoric*, atunci putem respinge că 'John nu este politicos' şi 'John este egoist' şi 'John este crud' sunt 'judecăţi de observaţie' (deşi, în cazuri particulare, s-ar putea cădea mai uşor de acord cu acestea decât, să spunem, că un obiect este violet). Dar asemenea judecăţi apar constant în lucrările unor oameni de ştiinţă, cum sunt *istoricii*. Este îndoielnic că descrierile din cadrul istoriei, psihologiei şi limbajului comun pot să evite cuvinte precum 'politicos', 'egoist' etc. (şi unde ar trebui delimitat domeniul neutru-valoric: 'încăpăţânat' este neutru valoric? 'furios' este neutru valoric? dar expresia 'îi răsuceşte mâna cu sălbăticie' este valoric neutră?). În orice caz, a identifica *raţionalitatea* cu raţionalitatea ştiinţifică, aşa cum a fost descrisă, ar însemna să nu se ţină seama de statutul judecăţilor de valoare; ar însemna să spunem că aceste judecăţi nu sunt raţional confirmabile pentru că *sunt* judecăţi de valoare, deoarece raţionalitatea a fost *definită* exclusiv în domeniul observaţiei neutre şi pentru inferenţe cu premise neutre valoric. Dar de ce să acceptăm o asemenea definiţie?

Valori, fapte şi cunoaştere

Am arătat în Capitolul 6 că orice fapt este valorizabil şi oricare dintre valorile noastre se regăseşte la nivelul faptelor deoarece *faptul* (sau adevărul) şi *raţionalitatea* sunt interdependente. Un fapt este ceva în care este raţional să se creadă, sau, mai precis, noţiunea unui fapt (sau judecată adevărată) este o idealizare a noţiunii de judecată în care este raţional să se creadă. 'Acceptabil raţional' şi 'adevărat' sunt noţiuni intercondiţionate. Am argumentat că a fi raţional presupune criterii de *relevanţă* şi criterii de acceptabilitate raţională şi că în criteriile de relevanţă sunt prezente valori. Decizia privind adevărul unei imagini a lumii (sau adevărul prin prisma cunoştinţelor actuale) şi *răspunsurile la întrebările relevante* (în măsura în care suntem în stare să le dăm un

răspuns) scoate la lumină angajamentul nostru într-un sistem de valori. Fără valori, nu există nici fapte.

Cel puţin indirect, modul în care criteriile de relevanţă presupun valori poate fi ilustrat prin examinarea celor mai simple judecăţi. Fie judecata 'Pisica se află pe covor'. Dacă cineva o formulează într-un context particular, atunci foloseşte resurse conceptuale – noţiunile de 'pisică', 'pe' şi 'covor' – care depind de o anumită cultură şi a căror prezenţă evidenţiază aspecte cu privire la interesele şi valorile acestei culturi. Avem categoria de 'pisică' deoarece acceptăm împărţirea lumii în *animale* şi *non-animale* şi suntem interesaţi asupra *speciei* căreia îi aparţine un anumit animal. Este *relevant* că o *pisică* se află pe covor şi nu un simplu *lucru*. Avem categoria de 'covor' deoarece considerăm importantă împărţirea non-animalelor în *artefacte* şi *non-artefacte* şi suntem interesaţi în *scopul* şi *natura* unui artefact. De aceea, este relevant că pisica se află pe un *covor* şi nu pe *altceva*. Avem categoria 'pe' deoarece suntem interesaţi în *relaţiile spaţiale*.

Să reţinem ce am observat: am luat în considerare cea mai banală judecată imaginabilă, 'pisica se află pe covor' şi am găsit că presupoziţiile sale includ semnificaţia unor categorii precum *animat*, *inanimat*, *scop* şi *spaţiu* într-un anumit context. Pentru o minte care nu consideră relevante aceste categorii, 'pisica se află pe covor' ar fi tot aşa de *iraţională* precum propoziţia 'numărul obiectelor hexagonale din această cameră este 76' rostită în timpul unei întâlniri între doi tineri îndrăgostiţi.

Nu numai aspectele cele mai generale privind sistemul de valori se manifestă în categoriile noastre (artefact, nume de specii, termen pentru o relaţie spaţială) ci, aşa cum am stabilit în Capitolul 6, atunci când se utilizează cuvinte clasificatoare ('politicos', 'egoist') sunt prezente valori mult mai specifice (cum ar fi sensibilitatea şi compasiunea). Subliniez încă o

dată, că criteriile de relevanţă pun în evidenţă întregul sistem de valori.

Importanţa discuţiei despre relevanţă pentru problema ridicată în capitolul precedent ('Care este valoarea raţionalităţii?') este imediată. Dacă 'raţionalitatea' este abilitatea (sau mai curând sistemul integrat de abilităţi) care ne permite să determinăm ce *probleme* sunt importante şi ce *soluţii* sunt justificate, atunci şi valoarea este prezentă. Dar nu este nevoie de nici un argument că o asemenea concepţie despre raţionalitate este tot atât de încărcată de valoare ca şi noţiunea de relevanţă.

S-ar putea obiecta că am pus împreună factori care ar trebui separaţi aşa încât am mascat un fel de truc. Faptul că am avut în vedere doi factori, acceptabilitatea raţională şi relevanţa, arată persistenţa unei dihotomii între fapte şi valori. O persoană raţională ar putea obiecta că nu se poate aserta nimic justificat în această privinţă deoarece ceea ce alege o persoană cu privire la *interesant*, *important* sau *relevant* ar putea fi determinat de caracterul său sau chiar de sănătatea mintală, nu de raţionalitatea *cognitivă*.

Acceptabilitatea şi relevanţa sunt totuşi interdependente în contextele reale. Utilizarea oricărui cuvânt – cum sunt 'bun', 'conştiincios', 'roşu' sau 'magnetic' – presupune o istorie, o tradiţie de observare, generalizare, practică şi teorie, dar şi activitatea de *interpretare* a acelei tradiţii şi de adaptare la noi contexte, extinderea şi critica ei. Tradiţiile ar putea fi interpretate variat dar un cuvânt nu poate fi folosit în afara tradiţiei căreia îi aparţine. A fi într-o tradiţie, afectează ceea ce este privit ca 'acceptabil raţional'. Dacă ar exista o metodă prin care să verificăm toate judecăţile indiferent de conceptele pe care le conţin, atunci ar putea fi susţinută distincţia între abilitatea de a verifica judecăţi şi cunoaşterea desăvârşită a

unei mulţimi *relevante* de concepte; dar am văzut deja că nu există nici un motiv pentru a accepta mitul metodei.

Teoria celor două componente

Intuiţiile noastre despre intenţionalitate par să fie în conflict; cu siguranţă, nici o teorie filosofică nu ar reuşi să le reconcilieze. Pe de o parte, nu este adevărat că noi nu judecăm *niciodată* scopurile ca raţionale sau neraţionale; pe de altă parte, când suntem confruntaţi cu un caz precum cel al ipoteticei 'raţionalităţi naziste', nu vedem cum să criticăm justificat un sistem de scopuri inteligent elaborat ca *iraţional* chiar dacă îl găsim respingător moral.

O sugestie pentru a rezolva o asemenea problemă este următoarea: presupunând o dihotomie strictă fapte/valori putem să justificăm condamnarea celor care sunt interesaţi numai de cunoaşterea numărului de fire de păr de pe capetele oamenilor ca iraţionali pe baza faptului că au o percepţie neadecvată cu privire la *fapte* (ce înseamnă 'adecvat' este o altă problemă). Naziştii sunt în dezacord cu noi în ce priveşte *valorile*, de aceea ei nu sunt consideraţi iraţionali, înscriindu-se între cazurile care pot fi investigate prin mijloacele sugerate de Bernanrd Williams. În particular, argumentul împotriva fetişismului metodei, că nu putem, fără circularitate, să respingem prin 'evidenţa observaţională' judecăţi descriptive precum 'John este politicos', poate fi susţinut prin propunerea unei teorii că vocabularul moral-descriptiv al limbajului obişnuit are simultan două 'componente de sens'. Una este componenta *factuală*; există *standarde* general acceptate pentru politicos şi 'John este politicos' conţine informaţia că John

îndeplineşte aceste standarde[1]. Dar există şi o componentă *emotivă* a sensului: 'John este politicos' conţine o 'atitudine-pro' cu privire la un aspect al conduitei lui John. Dintre acestea, se afirmă, numai componenta *factuală* a judecăţii 'John este pliticos' poate fi apreciată ca raţională; acceptarea componentei emotive, împărtăşirea 'atitudinii-pro' nu este nici raţională, nici iraţională.

Noţiunea de 'factual' depinde de perspectiva filosofilor asupra alcătuirii lumii. Pentru un filosof materialist, componenta factuală din sensul unei judecăţi constă într-o judecată exprimabilă în vocabularul *fizicii.* Dar apare o dificultate comparabilă cu aceea care afectează fenomenalismul.

Ne reamintim că fenomenalismul este doctrina care susţine că toate judecăţile cu sens sunt traductibile în judecăţi despre senzaţii. 'Fizicalismul' (de tipul materialismului pe care îl abordăm acum) este doctrina că sensul 'factual' al oricărei judecăţi poate fi tradus fără pierderi în limbajul fizicii. Şi această doctrină se dovedeşte a fi falsă.

Pentru a explica de ce, să considerăm, nu o judecată moral-descriptivă, ci una psihologică, să spunem, '*X* se

---

[1] Ce anume înseamnă 'standarde' în acest context rămâne o problemă deschisă. Dacă 'John este politicos' este un adevăr descriptiv (componenta sa factuală este adevărată) dacă şi numai dacă *cei mai mulţi vorbitori ar fi de acord că John este politicos*, atunci ar urma din *această* analiză a 'componentei factuale' că nu poate exista o persoană pe care o judecată observaţională ar clasifica-o *corect* drept 'politicoasă', deşi cei mai *mulţi* vorbitori nu ar fi *de acord.* O asemenea abordare a 'componentei factuale' ar fi asemenea cu afirmaţia că toate adevărurile (cel puţin despre 'standarde') trebuie să fie 'publice', dar de ce ar trebui cineva să creadă *aceasta* numai dacă are majoritatea? (cum am văzut în Capitolul 5, afirmaţia că *toate* adevărurile sunt publice se respinge singură).

gândeşte la Viena'. Este limpede că, chiar dacă există condiţii necesare şi suficiente exprimabile în termeni de stări cerebrale pentru ca un organism arbitrar să gândească la Viena, ar fi nevoie de o teorie neurologică inimaginabil de dezvoltată (sau poate funcţionalist-psihologică) pentru a le stabili. Condiţiile de adevăr ale acestei judecăţi sunt dependente de context, relative la interes şi vagi. Nu există nici o raţiune să se creadă că măcar 'în principiu' există o expresie *finită* a limbajului fizic, care (în orice lume fizică posibilă) este adevărată despre *X* dacă şi numai dacă *X* se gândeşte la Viena. Nu numai că ar putea fi *fals* că există un echivalent într-un limbaj fizic finit pentru judecata '*X* se gândeşte la Viena' din limbajul comun, ci dacă un asemenea echivalent *există*, echivalenţa ar fi bazată pe o teorie empirică sau pe un grup de teorii care nu sunt cunoscute (poate că sunt atât de complicate încât oamenii nu le vor *cunoaşte niciodată*) şi care, cu certitudine, nu aparţin *sensului* judecăţii '*X* se gândeşte la Viena'. Pe scurt, este fals că '*X* se gândeşte la Viena' *înseamnă* '*X* este într-o stare cerebrală aşa-şi-aşa (determinată fizic sau funcţional)'.

Ce se susţine pentru '*X* se gândeşte la Viena' trebuie susţinut pentru orice predicat din limbajul comun ale cărui condiţii de aplicare nu sunt guvernate de concepte fizice. '*X* este politicos', chiar '*X* este maro', '*X* este un cutremur' şi '*X* este o persoană' – nu sunt, nici ele, traductibile în limbajul unei teorii fizice. Asta înseamnă că, dacă există două componente pentru sensul judecăţii '*X* este politicos' atunci se poate descrie numai 'sensul factual' al judecăţii, care este adevărată dacă şi numai dacă *X* este *politicos*. O asemenea analiză ajunge să trivializeze noţiunea de 'componentă factuală'.

Când spunem că teoria celor două componente se prăbuşeşte, nu negăm că '*X* este politicos' are o anumită

forţă emotivă. Dar nu are întotdeauna. Cum am arătat în Capitolul 6, putem folosi judecata '*X* este politicos' pentru multe scopuri: pentru a evalua, a descrie, a face predicţii, a explica etc. *Modurile de folosire* a unei judecăţi trebuie circumscrise în aşa fel încât existenţa unei *judecăţi* precum '*X* este politicos' să nu fie negată.

Moore şi 'eroarea naturalistă'

Afirmaţia lui Weber că 'judecăţile de valoare' nu pot fi confirmate este sursa dihotomiei de astăzi dintre fapt şi valoare; dar această dihotomie este prezentă şi la Moore (contrar propriilor sale intenţii). Scriind pe vremea când Bertrand Russell şi John Maynard Keynes, împreună cu alţi membri ai grupului Bloomsbury, erau încă studenţi, Moore susţinea teza că binele este o proprietate non-naturală, respectiv, în afara ontologiei filosofice şi ştiinţei. Apărarea non-naturalismului de către Moore a avut consecinţe; studenţii săi au fost convinşi că există entităţi precum 'proprietăţile non-naturale' (deşi Russell, cel puţin, a pierdut convingerea) dar filosofii de mai târziu de orientare naturalistă simţeau că Moore a recurs la o *reductio ad absurdum* pentru teza că există entităţi precum proprietăţile de valoare. În anii 1930, Charles Stevenson şi pozitiviştii logici au dezvoltat 'teoria emotivă a eticii', adică teoria că '*X* este bun' înseamnă 'Eu îl aprob pe *X*, acţionează aşa de bine!' sau ceva de acest fel. Proprietăţile de valoare au început să fie respinse pe temeiuri epistemologice, dar şi pe temeiuri ontologice; după cum s-a exprimat nu de mult John

Mackie[2], existenţa unor entităţi precum *atitudinile valorice* este compatibilă cu ştiinţele naturale, dar *proprietăţile de valoare* nu sunt compatibile cu ştiinţele naturii. Proprietăţile de valoare, afirmă Mackie, sunt 'ontologic dubioase' – o proprietate în a cărei existenţă ştiinţifică oamenii instruiţi ar trebui să creadă discontinuu ar fi foarte misterioasă.

Moore argumentează că binele nu poate fi o proprietate fizică (sau 'naturală') deoarece, dacă 'binele' este o proprietate de tipul 'contribuie la maximizarea utilităţii totale' (sau orice altă proprietate naturală, fizică sau funcţională, prin care o putem substitui), atunci

> (1) 'această acţiune nu este bună cu toate că ea contribuie la maximizarea utilităţii totale'

este o judecată autocontradictorie (nu doar falsă).

Dar chiar un utilitarist nu ar accepta că (1) este *autocontradictorie*. Asta arată, după cum afirmă Moore, că, deşi a fi bun şi a contribui la maximizarea utilităţii totale ar putea fi proprietăţi *corelate*, ele nu pot fi proprietăţi *identice*.

Argumentul lui Moore se bazează pe o presupunere pe care eu şi mulţi alţi filosofi ai limbajului o respingem astăzi. Mai întâi de toate, ea neagă implicit că ar putea exista *proprietăţi identice sintetic*. Dar, cum am subliniat în Capitolul 4, aceasta ar contraveni unor descoperiri ştiinţifice precum aceea că mărimea *temperaturii* este aceeaşi cu mărimea *energiei cinetice moleculare medii*. (Unii ar putea utiliza argumentul lui

[2] *Ethics, Inventing Right and Wrong*, Penguin, 1977.

Moore pentru a 'arăta' că temperatura trebuie să fie o proprietate 'non-naturală'. Pentru aceştia, nu este contradicţie când spunem '*x* are temperatura *T* dar *x* nu are energia cinetică moleculară medie *E*', unde *E* este valoarea energiei cinetice moleculare care corespunde temperaturii *T*, chiar dacă o asemenea judecată este întotdeauna falsă din perspectivă empirică. Astfel, Moore ar trage concluzia că temperatura este numai *corelată* cu energia moleculară medie; cele două proprietăţi nu pot fi *identice*). De fapt, Moore confundă *proprietăţile* şi *conceptele*. Există o noţiune de proprietate pentru care faptul că două *concepte* sunt diferite (să spunem 'temperatura' şi 'energia cinetică moleculară medie') nu conduce la o diferenţă între *proprietăţile* corespunzătoare. (În cazul multor proprietăţi fizice fundamentale, se descoperă ceva despre *lume*, nu despre *concepte*). *Conceptul* 'bine' poate să nu fie sinonim cu vreun concept fizic (la urma urmelor, limbajul moral-descriptiv şi limbajul fizic sunt extrem de diferite, în sensul lui Goodman), dar de aici nu urmează că *a fi bun* nu este aceeaşi proprietate cu *a fi P*, pentru unele proprietăţi fizice (sau mai curând funcţionale) *P* corespunzătoare. În general, un termen învăţat ostensiv pentru o proprietate (de exemplu, 'temperatură înaltă') nu este sinonim cu definiţia teoretică a acelei proprietăţi; este nevoie de cercetare teoretică şi empirică, nu de analiză lingvistică pentru a determina ce este temperatura (şi, aşa cum unii filosofi ar putea să sugereze, ce este *bunătatea*), nu ajunge o reflectare asupra sensurilor.

Una dintre ideile care sunt discutate în filosofia limbajului de câţiva ani după ce am introdus

'proprietăţile sintetic identice', şi care lărgeşte şi iluminează problema avută în vedere, este teza lui Saul Kripke despre adevărul 'necesar metafizic', care trebuie învăţat *empiric*, 'adevărul epistemologic contingent'.[3] Observaţia lui Kripke, aplicată la cazul temperatură/energie cinetică, este că, dacă cineva descrie o lume posibilă logic în care oamenii au senzaţii de fierbinte şi rece, există obiecte care sunt simţite fierbinţi şi obiecte simţite reci şi în care aceste senzaţii de fierbinte şi rece sunt *explicate printr-un mecanism diferit decât energia cinetică moleculară medie*, atunci *nu* spunem că avem de-a face cu o lume posibilă în care *temperatura nu este energia moleculară medie.* Mai degrabă spunem că în lumea respectivă, *un mecanism, altul decât temperatura, face ca anumite obiecte să fie simţite calde sau reci.* Odată ce am adoptat 'judecata de identitate sintetică' potrivit căreia temperatura *este* energia cinetică moleculară medie (în lumea *efectivă*), nu mai există o lume posibilă în care temperatura nu este energia moleculară medie.

O judecată care este adevărată în orice lume posibilă este numită în mod tradiţional 'necesară'. O proprietate avută în orice lume posibilă este numită tradiţional 'esenţială'. În această terminologie tradiţională, Kripke spune că 'temperatura este energia cinetică moleculară medie' este un adevăr *necesar chiar dacă nu este cunoscut a priori.* Judecata este empirică dar necesară. Sau, pentru a spune în alte cuvinte, a fi energie cinetică moleculară medie este o proprietate esenţială a tempe-

[3] Kripke S., *Naming and Necessity*, Harvard, 1980 (conferinţe susţinute iniţial la Princeton, în 1970).

raturii. Am descoperit *esența* temperaturii prin investigare *empirică*. Aceste idei ale lui Kripke au avut un impact deosebit în filosofia limbajului, metafizică și filosofia matematicii; aplicate la argumentul lui Moore ele sunt hotărâtoare. Moore argumentează din faptul că (1) *poate fi numai contingent fals* că *P* (pentru o proprietate naturală corespunzătoare) nu ar putea fi o proprietate esențială pentru bunătate; adică tocmai ceea ce interzice noua teorie a necesității. Din faptul că (1) nu este autocontradictorie se poate infera valid că 'bine' nu este sinonim cu 'a contribui la utilitatea maximă' (nu este sinonim cu *P*, pentru orice termen *P* în versiunea fizicalistă a lumii). Din această asinonimie a *cuvintelor* nu urmează nimic cu privire la non-identitatea *proprietăților*. Nu urmează nimic despre esența binelui.

Ruth Ana Putnam a arătat că un alt argument potrivit căruia binele nu poate fi o proprietate naturală nu funcționează[4]. Este vorba de argumentul că '*X* este bun' are o 'forță emotivă', 'expresie a atitudinii-pro' etc.

Acest argument este eronat deoarece multe predicate descriptive naturale *dobândesc* o forță emotivă. În cultura noastră, 'îi curge supa pe cămașă' are o forță negativă puternică, deși fraza este, literal, o *descriere*. Orice cuvânt care stă pentru cineva ca valoare culturală (sau non-valoare) va tinde să dobândească o forță emotivă.

---

[4] 'Remarks on Wittgenstein's Lecture on Ethics', Haller și alții (eds.), *Language, Logic and Philosophy, Proceedings of the 4th Intern. Wittgenstein Symposium*, Vienna, 1980.

Cuvântul 'bine', în sens moral, se aplică la multe lucruri. Unele dintre acestea – stările mentale bune, de pildă – pot fi evaluate natural: pot fi parte a conţinutului chiar a stării mentale valorizate. De aici nu decurge conversa: că orice stare mentală valorizată natural în acest sens este bună. Să presupunem că 'binele' ar fi definit aşa încât lucrurile care sunt apreciate în mod natural şi cele pentru care nu există motive pentru a fi respinse (după cum există motivaţie pentru a respinge stările mentale induse prin droguri) sunt considerate 'bune'.

Atunci s-ar crede că judecata potrivit căreia ceva este bun are o forţă emoţională pozitivă *datorită naturii proprietăţii*. Chiar pentru cei care nu sunt consecvenţialişti (aceştia cred că *orice* care are suficiente *consecinţe* bune este bun), nu există îndoială că motivul cel mai potrivit pentru a numi o acţiune bună este că are consecinţe bune, între care s-ar putea afla stări sau situaţii care sunt apreciate natural; din nou, adevărata natură a proprietăţii explică de ce descrierea ajunge să aibă o forţă emotivă 'pro'.

Mackie apără afirmaţia sa că binele este ontologic neclar introducând ca premisă presupunerea că nu putem *cunoaşte* că un lucru este bun fără a avea o atitudine 'pro'. Asta înseamnă să presupunem emotivitatea ca dovadă pentru emotivitate. Diavolul din iad este înfăţişat adesea prin utilizarea lui 'bine' cu o forţă emotivă *negativă* ('El are o tendinţă deplorabilă spre binele moral' s-ar putea spune); împotriva lui Mackie, nu consider că asemenea utilizări ale limbii sunt improprii sau generatoare de contradicţii. Oare nu auzim oamenii spunând 'Ştiu că

este un lucru rău dar nu-mi pasă'? După cum a arătat Philippa Foot, cineva ar putea înceta să apeleze la moralitate spunând 'Eu nu am de gând să fiu un om bun'. Aceste enunţuri arată că, de vreme ce există o diferenţă între *folosirea descriptivă* a limbajului şi *folosirea prescriptivă* sau *folosirea imperativă*, apare posibilitatea ca aici să nu fie implicat numai vocabularul. Cuvintele 'descriptive' pot fi folosite pentru a lăuda sau blama ('Lui îi curg balele pe cravată') şi cuvintele 'evaluative' pot fi folosite pentru a descrie şi explica. (Să ne oprim asupra următorului dialog: 'John trebuie să fie un om deosebit de bun pentru a face asta'. 'Nu, el nu a fost niciodată un exemplu moral, ba dimpotrivă; dar capacitatea de sacrificiu de sine nu i se poate pune la îndoială'. Aici limbajul *moral* este folosit într-o funcţie *explicativă*). Pentru a repeta ceea ce sublinia Ruth Anna Putnam, din faptul că 'este bine' se foloseşte în sens imperativ nu urmează că binele nu poate fi o proprietate.

D-na profesor Putnam sublinia că în argumentul lui Mackie există ceva *corect*, de bună seamă. Unele expresii morale au în mod cert o orientare spre acţiune. 'Ar trebui', 'se cuvine', 'corect' etc. sunt exemple de cuvinte 'care ghidează acţiunea'. Problema 'este/se cuvine' nu este aceeaşi cu problema 'fapt/valoare', aşa cum a arătat ea. 'Nu sunt ce ar trebui' este mai nedeterminat decât 'Eu nu sunt cum ar trebui să fie un om bun'.

Mackie arată că nici o proprietate fizică nu are o conexiune orientativă cu acţiunile şi trage concluzia că 'a fi un lucru corect' etc., sunt 'ontologic neclare'. Dar, (dincolo de dependenţa faţă de presupunerea că versiunea

fizicalistă a lumii este teoria adevărului unic), acest argument este exagerat deoarece unele predicate *epistemice* (cum ar fi 'raţional acceptabil', 'justificat' etc.) sunt orientative în ceea ce priveşte acţiunea (într-un sens larg, acceptarea unei judecăţi poate fi privită ca acţiune). S-ar putea spune '*X* este un lucru bun de făcut' şi 'Este o bine să consideri că *Y* este evident' fără angajamentul de a face *X* sau să de a accepta *Y*; dar cine spune '*X* este reacţia corectă faţă de această situaţie' sau 'A crede *Y* este complet justificat', acesta orientează acţiunea către a face *X* sau a crede *Y*. 'Justificat' (în cazul convingerilor) orientează acţiunile tot aşa de mult ca şi 'corect' în sens moral.

Dacă acceptăm argumentul lui Mackie şi tragem concluzia că nu există vreo proprietate ca *justificat*, ci numai 'atitudini de justificare', atunci ajungem la relativism total. Înainte de a cădea într-o asemenea catastrofă, ar trebui să ne oprim să vedem *de ce* predicatele care orientează acţiunea par ontologic neclare pentru un fizicalist.

În Capitolul 2 am argumentat că şi *referinţa* trebuie privită ca 'ontologic obscură' de către un fizicalist. Dacă există numai proprietăţile şi relaţiile fizice, atunci referinţa, pentru a exista, trebuie să fie o relaţie fizică – dar în acest caz ar apare problema abundenţei candidaţilor pentru o asemenea relaţie. Există un număr infinit de relaţii de referinţă admisibile (şi toate sunt fizice sau cel puţin naturale, în cazul în care considerăm *teoria mulţimilor* ca parte a viziunii naturaliste asupra lumii). Dacă una dintre acestea ar fi relaţia de referinţă, atunci *acest* fapt ar fi *metafizic* straniu pentru că trebuie

să determine o anumită *neutralitate*, o anumită *neangajare mentală* în versiunea noastră despre natură. Natura este presupusă a nu avea interese, intenţii sau puncte de vedere. Dacă aşa stau lucrurile, cum ar *putea* fi selectată o singură relaţie de referinţă admisibilă metafizic?

Aceeaşi nonintenţionalitate a naturii face ambigue predicatele care orientează acţiuni, cum sunt 'este corect', 'este justificat să credem'. Dacă o proprietate fizică *P* ar fi corectitudinea morală sau justificarea epistemologică, aceasta ar fi 'ambiguă' - din aceleaşi motive pentru care referinţa este 'ambiguă' din perspectivă fizicalistă. Aşa s-ar întâmpla dacă natura însăşi ar avea valori, în cazul moral, sau intenţii, în cazul semantic.

Dintr-un asemenea motiv, cred că Moore avea dreptate (chiar dacă argumentul său nu este acceptabil) când susţinea că 'bine', 'corect' (ca şi 'justificat', 'se referă' şi 'adevărat') nu sunt identice cu proprietăţi şi relaţii fizice. De aici nu rezultă că bunătatea, dreptatea, justificarea epistemică, referinţa şi adevărul nu există, ci numai că naturalismul monar (sau 'fizicalismul) este o filosofie neadecvată.

Din nou, 'nazismul raţional'

Mai înainte, eram preocupaţi cum să argumentăm că ipoteticii 'naziţti perfect raţionali' aveau scopuri iraţionale. Pentru a răspunde, poate că întrebarea privind raţionalitatea *nazismului* (privit ca perspectivă asupra lumii) ar trebui înlocuită cu întrebarea privind

raţionalitatea *scopurilor* nazismului. Dacă nu există nici o finalitate a nazismului la care să ne putem raporta, atunci, putem spune 'Karl are scopuri iraţionale' cu totul la întâmplare. Mai întâi, urmărim să afirmăm despre Karl că are scopuri *monstruoase*, nu atât iraţionale.

Dar, dacă intenţionăm să vorbim despre raţionalitatea lui Karl, trebuie să investigăm iraţionalitatea *convingerilor şi argumentelor*, nu a scopurilor.

Să presupunem mai întâi despre personajul nostru că afirmă despre scopurile naziştilor că sunt *drepte* şi *bune* (aşa cum au făcut naziştii). Atunci, *de fapt*, el spune prostii. El va relua toate tipurile de propoziţii 'factuale' false, cum ar fi că democraţia este rodul 'conspiraţiei evreieşti' şi va considera propoziţii morale (de pildă, că arienii au *datoria* să subjuge rasele non-ariene ca 'rasă dominantă') pentru care nu are argumente potrivite. Raportez noţiunea de 'argument bun' la discursul moral *intern*; dar aceasta devine o noţiune specifică, dacă naziştii încearcă să se justifice în *interiorul* discursului moral intern.

Să presupunem, pe de altă parte, că naziştii *resping* noţiunile morale obişnuite (aşa cum ar face superbenthamiştii). O cultură care respinge noţiunile morale obişnuite sau le înlocuieşte cu noţiuni derivate din alte ideologii şi morale, pierde abilitatea de a *descrie* relaţiile interpersonale, evenimentele sociale şi politice *din perspectiva noastră de astăzi*. Bineînţeles, dacă o ideologie şi o morală diferite sunt *superioare* ideologiei şi moralei noastre actuale, atunci substituirea poate fi bună şi înţeleaptă; dar dacă acestea sunt *rele* îndeosebi dacă sunt deformate şi monstruoase, atunci rezultatul va fi o

reprezentare inadecvată şi respingătoare a faptelor interpersonale şi sociale. De bună seamă, acestea sunt judecăţi de valoare; dar am argumentat că alegerea unei scheme conceptuale reflectă *cu necesitate* judecăţi de valoare şi alegerea unei scheme conceptuale este ceea ce reprezintă raţionalitatea *cognitivă*.

Nazistul nu pierde abilitatea de a folosi vocabularul nostru moral-descriptiv prezent, chiar dacă noţiunile vechi rămân cumva în mintea sa (aşa cum unii învăţaţi, poate, sunt încă familiari cu şi sunt în stare să folosească noţiunea medievală de 'cavalerism'), în schimb, noţiunile noastre moral-descriptive, precum 'politicos', 'milos', 'drept', nu le-a folosit în decursul vieţii sale: ele nu figurează ca atare în construcţia proprie a lumii.

Din nou, aş dori să subliniez că eu nu afirm că *răul* din a fi nazist constă în convingeri deformate şi iraţionale, ci în *a face*. Nazistul este rău *şi* are, pe deasupra, o viziune iraţională asupra lumii. Aceste două fapte despre nazism sunt legate şi intercorelate; dar asta nu înseamnă că nazismul este rău în primul rând *pentru că* are o viziune iraţională asupra lumii în sensul că iraţionalitatea *constituie* răul. Neîndoielnic, *există* un sens în care putem vorbi de *scopuri* ca fiind raţionale sau iraţionale, după câte cred: sunt numite pe drept cuvânt 'scopuri iraţionale' scopurile pe care, dacă cineva le acceptă şi le urmează, este nevoit să le justifice prin argumente false (dacă cineva vrea să le justifice în interiorul schemei conceptuale normale) sau va fi nevoit să adopte o schemă alternativă pentru a reprezenta faptele obişnuite moral-descriptive (de pildă, faptul că o persoană este milostivă) care este iraţională. Există o

legătură, la urma urmei, între folosirea unei scheme conceptuale raţionale în descrierea şi perceperea faptelor relevante moral şi anumite tipuri generale de scopuri.

'Dar ce se întâmplă dacă nazismul nu oferă altă raţiune pentru a fi nazist decât "aşa simt eu să acţionez"?' Aceasta este o întrebare firească, iar răspunsul este: într-un asemenea caz, comportamentul nazist, pe lângă că este rău, este şi complet *arbitrar*. Să reţinem că 'arbitrar' este unul dintre cuvintele pe care le consider 'moral-descriptive', este un cuvânt care poate fi folosit, fără a-i modifica denotarea, pentru a evalua (în acest caz, pentru a blama), pentru a descrie ('John s-a decis cu totul arbitrar să-şi părăsească slujba'), pentru a explica (sau a indica faptul că nici o explicaţie de un anumit fel nu poate fi dată) etc. Într-adevăr, când am spus că decizia lui Karl de a fi nazist (în cazul descris) ar fi complet arbitrară, mai întâi am *descris*, nu am evaluat. Multe lucruri pe care le fac sunt, în mod literal, arbitrare – de pildă, când aleg o rută prin campus mai degrabă decât alta; dar de aici nu rezultă că este ceva rău cu acţiunile mele. Chiar dacă fac ceva important în mod arbitrar – cum ar fi alegerea unei slujbe – dacă nu am responsabilităţi familiale etc., poate fi, pur şi simplu, *dreptul* meu. Dar dacă acţiunea este una ca *cere o justificare* atunci, a o efectua arbitrar şi fără justificare înseamnă expunerea la un blam legitim. A lua o decizie care afectează vieţile altora (poate chiar propria viaţă) fără nici o justificare, în mod arbitrar şi fără a implica voinţa (alt cuvânt moral-descriptiv!) este un exemplu paradigmatic de iraţionalitate, ba chiar, nu doar iraţionalitate, ci şi perversitate.

Am început discuţia în Capitolul 7 cu afirmaţia lui Bentham că 'fără prejudecăţi', jocul de pushpin este tot atât de bun precum 'artele şi ştiinţele muzicii şi poeziei'. Din perspectiva lui Bentham, singurul motiv pentru care poezia este mai bună decât pushpin, în ultimă instanţă, este faptul brut că poezia oferă o satisfacţie mai mare decât pushpin (sau satisface mai mulţi oameni, ori ambele). Există, în fond, două lucruri eronate în această concepţie: un lucru greşit este 'satisfacţia' (sau 'interesul propriu') care nu poate fi scop în sine fără a *presupune* alte scopuri. Dacă nu am alt scop decât 'bunăstarea' proprie, atunci 'bunăstarea' nu are nici un sens, ceea ce ne duce înapoi la episcopul Butler. Mai important, unele *satisfacţii* sunt mai bune sau mai 'nobile' decât altele şi se poate argumenta de ce. Poezia şi muzica oferă mângâiere, îmbogăţesc sensibilitatea, oferă posibilităţi aparte de exprimare pentru mulţi oameni, inclusiv pentru oamenii cei mai dotaţi. Numind 'prejudecată' motivele pentru a valoriza anumite satisfacţii în raport cu altele se ajunge la teoria celor două componente *şi* la teza că proprietăţile valorice sunt 'ontologic obscure'. Bentham operează cu un model al 'faptelor neutre' şi al 'prejudecăţilor' arbitrare. Într-adevăr, numind preferinţa pentru poezie 'prejudecată', Bentham sugerează că orice preferinţă pentru un *tip* de satisfacţie în raport cu altul este arbitrară. Dar asta este fals dacă ţinem seama de locul din schema conceptuală a noţiunii de preferinţă 'arbitrară' şi fără sens dacă 'arbitrar' este scos din schema căreia îi aparţine. (În mod asemănător, judecata că preferinţa poeziei în relaţie cu pushpin este o prejudecată este *falsă literal*.) S-a sugerat că, într-un fel, este legitim ontologic

să se admită că există asemenea lucruri, cum sunt *satisfacţiile*, dar nu este legitim ontologic să se admită îmbogăţirea sensibilităţii, obţinută prin metafore, moduri de expresie şi autorealizare etc. Ideea că valorile nu sunt elemente ale lumii şi ideea că 'judecăţile de valoare' sunt expresii ale 'prejudecăţilor' sunt două feţe ale aceleiaşi monede.

Am cercetat problema dacă 'judecăţile de valoare' pot fi susţinute raţional. Am trecut în revistă diferite răspunsuri negative şi presupoziţii filosofice îndoielnice: că raţionalitatea în sine este bună numai pentru 'predicţii' sau numai pentru 'consens' sau că există o singură metodă pentru a dobândi adevărul (unde, uneori, singurele criterii pentru 'adevăr' sunt date de predicţii şi consens) sau că judecăţile de valoare au 'două componente de sens' sau că proprietăţile valorice sunt ontologic 'obscure'. Poziţia pe care am apărat-o este că orice alegere a vreunei scheme conceptuale presupune valorile, iar alegerea unei scheme pentru a descrie relaţiile interpersonale şi faptele sociale, fără a menţiona ideile despre propriul plan de viaţă, implică între alte lucruri şi valori *morale*. Nu se poate alege o asemenea schemă numai printr-o simplă reproducere a 'faptelor', deoarece nici o schemă conceptuală nu este o simplă 'copie' a lumii. Chiar şi noţiunea de adevăr depinde de standardele proprii de acceptabilitate raţională, iar aceasta presupune apelul la valori. Poate prea schematic, am arătat că teoria adevărului presupune teoria raţionalităţii, care, la rândul ei, presupune o teorie a binelui.

'Teoria binelui', totuşi, nu este numai programatică, ci şi dependentă de presupunerile privind natura

umană, societatea, universul (incluzând presupuneri teologice şi metafizice). Teoria binelui trebuie mereu adaptată cunoaşterii în creştere şi schimbărilor din tabloul asupra lumii.

A devenit clar că în concepţia pe care o apăr aici nu apare nimic de felul unui 'fundament'. De aceea, unii ar putea fi îngrijoraţi: oare nu suntem prea aproape de perspectiva în care nu există diferenţă între 'justificat' şi 'justificat după propriile criterii' (relativism) sau chiar 'justificat numai potrivit criteriilor mele' (o specie de solipsism)?

Într-adevăr, ne aflăm în poziţia solipsistului dacă situându-ne în afara sistemului conceptual căruia îi aparţine conceptul de raţionalitate pretindem să oferim o noţiune mai 'raţională' a raţionalităţii! (Mulţi gânditori au căzut în eroarea lui Nietzsche spunându-ne că ei posedă cheia unei moralităţi mai bune decât cea tradiţională; în fiecare caz ei au ajuns la monstruozităţi, deoarece au scos în mod arbitrar anumite valori în afara contextului lor în vreme ce altele au fost ignorate). Putem doar să sperăm într-o *concepţie* mai raţională privind raţionalitatea dacă operăm în *interiorul* tradiţiei noastre (cu ecourile sale în agora grecilor, în gândirea lui Newton etc., în cazul raţionalităţii, şi cu ecourile sale în scriptură, în filosofi, în revoluţiile democratice etc., în cazul moralităţii); dar asta nu înseamnă de loc că totul este pe de-a-ntregul raţional şi bun în ce priveşte concepţiile pe care le avem acum. Nu suntem prinşi în capcana solipsismului, ci invitaţi să ne angajăm într-un dialog adevărat; fiecare îmbină responsabilităţile colective şi individuale.

Oare acest dialog are un ideal spre care tinde? Există o concepţie *adevărată* asupra raţionalităţii, o moralitate *adevărată*, chiar dacă fiecare avem propriile *concepţii* cu privire la acestea? Aici filosofii se împart, ca în orice altă privinţă. Richard Rorty, în Adresa Prezidenţială[5] către Asociaţia Americană de Filosofie, optează pentru punctul de vedere că există numai dialogul; nu există nici un ideal ca finalitate şi nici nu este nevoie. Dar cum anume aserţiunea că 'există numai dialogul' diferă de relativismul care se autorespinge pe care l-am abordat în Capitolul 5? Faptul că discutăm asupra a diferite concepţii ca fiind concepţii diferite ale *raţionalităţii* ne îndreaptă spre un *Grenzbegriff*, un concept-limită al adevărului ideal.

[5] 'Pragmatism, Relativism and Irrationalism', *Proceedings and Addresses of the American Philosophical Association*, August 1980. Vezi, de asemenea, R. Rorty, *Philosophy and the Mirror of Nature*, Princeton University Press, 1979.

Anexă

Aici este prezentată teorema la care m-am referit în Capitolul 2.

Teoremă. Fie L un limbaj cu predicatele F_1, F_2, ..., F_k (nu trebuie să fie monadice). Fie I o interpretare, prin care se atribuie câte o intensiune fiecărui predicat din L. Atunci, dacă I nu este trivială, în sensul că există cel puţin un predicat având o extensiune care nu este nici vidă nici universală în cel puţin o lume posibilă, există cel puţin o a doua interpretare J diferită de I, dar care face ca aceleaşi propoziţii să fie adevărate în orice lume posibilă ca şi I.

Demonstraţie. Fie W_1, W_2, ..., toate lumile posibile, într-o anumită ordine bine definită şi fie U_i o mulţime de individuali posibili care există în lumea W_i. Fie R_{ij} o

mulţime care este extensiunea predicatului F_i în lumea posibilă W_j în conformitate cu I (dacă F_{ij} nu este monadic, atunci R_{ij} este o mulţime de n-tuple, unde n_i este numărul de argumente pe care le acceptă F_i). Sistemul $< U_j; R_{ij} (i = 1,2,..., k) >$ este 'modelul intenţionat' a lui L în lumea W_j, relativ la I (U_j este universul de discurs a lui L în lumea W_j şi (pentru $i = 1, 2, ..., k$) R_{ij} este extensiunea predicatului F_i în W_j.

Cel puţin un predicat, să spunem, F_u, cu o extensiune R_{uj} care nu este nici vidă nici universală (U_j), selectează o permutare P_j a elementelor din U_j astfel încât $P_j(R_{uj}) \neq R_{uj}$. Astfel, fie P_j identitatea. Dacă P_j este o permutare, sistemul $< U_j; P_j(R_{ij}) (i = 1, 2, ..., k) >$ este izomorf cu $< U_j; R_{ij} (i = 1, 2, ..., k) >$ aşa încât există un model pentru aceleaşi propoziţii în L (cum ar fi propoziţiile din L care sunt adevărate sub I în W_j).

Fie J interpretarea lui L care asociază predicatului F_i ($i = 1, 2, ..., k$) următoarea intensiune: funcţia $f_i(W)$ a cărei valoare în orice lume posibilă W_j este $P_j(R_{ij})$. Cu alte cuvinte, extensiunea lui F_i în fiecare W_j sub interpretarea J este definită ca fiind $P_j(R_{ij})$. Dacă $< U_j; P_j(R_{ij}) (i = 1, 2, ..., k) >$ este un model pentru aceeaşi mulţime de propoziţii ca şi $< U_j; R_{ij} (i = 1, 2,..., k) >$ (prin izomorfism), atunci aceleaşi propoziţii sunt adevărate în orice lume posibilă sub J ca sub I, deşi J diferă de I în orice lume posibilă în care cel puţin un predicat are o extensiune netrivială. q.e.d.

Comentariu: dacă, într-o lume dată W_j, există două mulţimi disjuncte care sunt extensiuni ale predicatelor din L în W_j sub I – să spunem o mulţime de pisici şi o mulţime de câini – atunci, dacă există mai mulţi câini decât pisici (respectiv, cel puţin atâtea pisici câţi câini) putem lua în considerare orice mulţime de câini de aceeaşi mărime cu

mulţimea de pisici (respectiv, tot atâtea pisici câţi câini) şi să alegem P_j care selectează mulţimea aleasă de câini (respectiv mulţimea aleasă de pisici) şi viceversa; aceasta este necesar pentru ca sub J extensiunea primului predicat – cel a cărui extensiune sub I este mulţimea de pisici – este o mulţime de câini sub J în W_j, sau extensiunea celui de-al doilea predicat – iar dacă extensiunea sub I este o mulţime de câini – sub J să fie o mulţime de pisici în W_j.

Al doilea comentariu: dacă există obiecte – să spunem, 'senzaţii' – care am dori să *nu* fie permutate, deoarece cu privire la predicate, acestea sunt într-un anume sens 'absolute', se poate stipula că P_j acţionează drept permutare identică în ceea ce priveşte asemenea obiecte. Apare, astfel, o restricţie pentru orice predicat din L pentru asemenea obiecte privilegiate sub ambele interpretări, în fiecare lume.

Al treilea comentariu: dacă propoziţiile primesc condiţii de echivalenţă logică în ceea ce priveşte adevărul sub ambele interpretări, urmează că în semantica standard a lumilor posibile, *condiţionalii contrafactuali* se conservă.

taias

raţiune adevăr şi istorie

Lucrarea de faţă înfăţişează stadiul ideilor filosofice ale lui Putnam la începutul deceniului IX asupra problemei referinţei, problemei raportului minte-corp, problemei realismului şi relativismului, dar şi asupra altor teme precum raportul fapt-valoare, raţionalitatea şi filosofia istoriei etc.

Filosofia ştiinţei este abordată de Putnam îndeosebi din perspectiva disputei dintre realism şi empirism. Putnam respinge *realismul metafizic*, care este legat de teoria adevărului corespondenţă, argumentând că nu există o perspectivă absolută asupra realităţii şi adoptă *realismul intern*, potrivit căruia limbajul şi realitatea se află în interacţiune, astfel încât limbajul contribuie la construirea realităţii, dar şi realitatea are rolul său în construirea limbajului.

Ionel Narița

ISBN 973-31-2259-9